

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра дифференциальных уравнений и системного анализа

Аннотация к дипломной работе

**ПРОГНОЗИРОВАНИЕ ВЫБОРА КЛИЕНТОМ ОТЕЛЯ С ПОМОЩЬЮ
МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ**

Мурашко Александра Сергеевна

Научный руководитель:
кандидат физ.-мат. наук,
доцент Л.Л. Голубева

2017

В дипломной работе 78 страниц, 13 рисунков, 8 таблиц, 34 источника, 2 приложения.

Ключевые слова: МАШИННОЕ ОБУЧЕНИЕ, АНАЛИЗ ДАННЫХ, БОЛЬШИЕ ДАННЫЕ, АЛГОРИТМЫ, КЛАССИФИКАЦИЯ, ДЕРЕВЬЯ ПРИНЯТИЯ РЕШЕНИЙ, АНСАМБЛЕВЫЕ МЕТОДЫ ОБУЧЕНИЯ, ОБЛАЧНЫЕ ВЫЧИСЛЕНИЯ.

В дипломной работе изучаются методы машинного обучения применительно к задаче мультиклассификации.

Целью дипломной работы является исследование проблемы выявления внутренних закономерностей больших объемов данных методами машинного обучения. Для достижения поставленной цели решены следующие задачи:

- проведена предварительная обработка данных
- построены различные модели алгоритмов машинного обучения
- проведен анализ влияния параметров модели на эффективность обучения.

В дипломной работе получены следующие результаты:

1. Описаны использованные в работе техники по работе с данными, алгоритмы машинного обучения и технологии, с помощью которых осуществлялась работа
2. Построены модели с использованием различных алгоритмов (решающие деревья, Random Forest, Gradient Boosted Trees, Stacking)
3. Приведено сравнение производительности и точности построенных моделей

Дипломная работа носит практический характер. Ее результаты могут быть использованы для исследований, связанных с большими выборками данных.

Обоснованность и достоверность полученных результатов обусловлена согласованностью с результатами, известными ранее для частных случаев.

Дипломная работа выполнена автором самостоятельно.

Diploma thesis consists of 78 pages, 13 figures, 8 tables, 34 references and 2 appendices.

Key words: MACHINE LEARNING, DATA ANALYSIS, BIG DATA, ALGORITHMS, CLASSIFICATION, DECISION TREES, ENSEMBLE LEARNING METHODS, CLOUD COMPUTING.

The main topic of this diploma thesis is study of different methods of machine learning for multiclassification.

The purpose of this study is to investigate the problem of detecting patterns in big data using machine learning methods. To reach the desired goal the following problems were solved:

- preprocessing of the data performed
- different machine learning models were built
- the influence of the model's parameters on the performance and accuracy was analyzed

The results of the diploma thesis:

1. Techniques and technologies used for data processing, machine learning algorithms were described
2. Models based on different algorithms (decision tree, Random Forest, Gradient Boosted Trees, Stacking) were built
3. Performance and accuracy of the built models compared.

The diploma thesis is a practical one. Its results can be used for researches that include processing of big data. The validity and reliability of the obtained results justified by the coherence with the known results for special cases.

The thesis project was done solely by the author.