

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра многопроцессорных систем и сетей

Аннотация к дипломной работе

**«ОБРАБОТКА БОЛЬШИХ МАССИВОВ ДАННЫХ С
ПРИМЕНЕНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ»**

Барашко Мария Андреевна

Научный руководитель – старший преподаватель О.М. Кондратьева

2016

Реферат

Дипломная работа, 50 страниц, 12 рисунков, 14 источников, 12 формул.

БОЛЬШИЕ ДАННЫЕ, АНАЛИЗ ДАННЫХ, ИНТЕЛЛЕКТУАЛЬНЫЙ
АНАЛИЗ ДАННЫХ, МАШИННОЕ ОБУЧЕНИЕ, КЛАССИФИКАЦИЯ,
TWITTER

Объект исследования — применимость методов машинного обучения для классификации текстовых документов.

Цель работы — исследовать методы машинного обучения для осуществления бинарной классификации больших массивов данных.

Результат работы — программный инструмент, позволяющий получать потоковые данные из микроблоггинга Twitter, преобразовывать данные для осуществления классификации, строить классификационную модель при помощи различных алгоритмов классификации, анализировать полученные результаты.

Abstract

Graduation thesis, 50 pages, 12 figures, 14 references, 12 formulas.

**BIG DATA, DATA ANALYSIS, DATA MINING, MACHINE LEARNING,
CLASSIFICATION, TWITTER**

Object of research — applicability of the machine learning methods in the area of text documents classification.

Purpose — research of the machine learning methods for implementation of the binary classification on the large amounts of data.

Results — an application that can be used to receive streaming data from microblogging Twitter, to convert data for classification, to build a classification model using different classification algorithms for the result analysis.