

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
Кафедра технологий программирования**

Аннотация к дипломной работе

**«РЕАЛИЗАЦИЯ ПРИЛОЖЕНИЯ ДЛЯ АВТОМАТИЗАЦИИ
ЭТАПОВ НАСТРОЙКИ МОДЕЛЕЙ МАШИННОГО
ОБУЧЕНИЯ»**

Ястребова Вероника Алексеевна

Научный руководитель – кандидат физико-математических наук В.В. Мушко

Минск, 2020

РЕФЕРАТ

Дипломная работа, 60 с., 45 рис., 6 формул, 14 источников, 6 приложений.

Ключевые слова: БОЛЬШИЕ ДАННЫЕ, МАШИННОЕ ОБУЧЕНИЕ, АВТОМАТИЗАЦИЯ, JAVA, APACHE SPARK.

Объект исследования – процесс построения модели машинного обучения.

Предмет исследования – особенности, проблемы различных этапов при построении модели.

Цель работы – исследовать особенности, рассмотреть возможные проблемы и решения, возникающие при построении модели машинного обучения.

Задачи:

1. обзор проблемы больших данных и способов ее решения;
2. обзор основных этапов построения моделей машинного обучения;
3. реализация приложения для отображения выделенных этапов и их настройки, позволяющего получить обученную модель машинного обучения и оценить ее точность;
4. анализ функционирования приложения на наборах реальных данных

Методы исследования – аналитическое сравнение.

Полученные результаты:

1. подготовлен аналитический обзор этапов, выделяемых при построении моделей машинного обучения;
2. рассмотрены наиболее популярные методы в рамках выделенных этапов;
3. выявлены достоинства и недостатки, а также найдены основные ограничения, при которых могут использоваться рассмотренные методы;
4. реализовано приложение и продемонстрирована его работа на наборе реальных данных.

Область применения – машинное обучение, большие данные.

РЭФЕРАТ

Дыпломная праца, 60 с., 45 рыс., 6 формул, 14 крыніц, 6 прыкладанняў.

Ключавыя словы: ВЯЛІКІЯ ДАДЗЕНЫЯ, МАШЫННАЕ НАВУЧАННЕ, АУТАМАТЫЗАЦЫЯ, JAVA, APACHE SPARK.

Аб'ект даследавання – працэс пабудовы мадэлі машыннага навучання.

Прадмет даследавання – асаблівасці, праблемы розных этапаў пры пабудове мадэлі.

Мэта працы – даследаваць асаблівасці, разгледзець магчымыя праблемы і рашэнні, якія ўзнікаюць пры пабудове мадэлі машыннага навучання.

Задачы:

1. агляд праблемы вялікіх дадзеных і спосабаў яе вырашэння;
2. агляд асноўных этапаў пабудовы мадэляў машыннага навучання;
3. рэалізацыя прыкладання для адлюстравання выдзеленых этапаў і іх налады, які дазваляе атрымаць навучанаю мадэль машыннага навучання і ацаніць яе дакладнасць;
4. аналіз функцыянавання прыкладання на наборах рэальных дадзеных.

Метады даследавання – аналітычнае параўнанне.

Атрыманыя вынікі:

1. падрыхтаваны аналітычны агляд этапаў, што выдзяляюцца пры пабудове мадэляў машыннага навучання;
2. разгледжаны найбольш папулярныя метады ў рамках выдзеленых этапаў;
3. выяўленыя вартасці і недахопы, а таксама знойдзеныя асноўныя абмежаванні, пры якіх могуць выкарыстоўвацца разгледжаныя метады;
4. рэалізавана прыкладанне і прадэманстравана яго праца на наборах рэальных дадзеных.

Вобласць ужывання - машыннае навучанне, вялікія дадзеныя.

ABSTRACT

Diploma, 60 p., 45 images, 6 equations, 14 sources, 6 appendixes.

Key words: BIG DATA, MACHINE LEARNING, AUTOMATIZATION, JAVA, APACHE SPARK.

Object of research – the process of building a machine learning model.

Subject of research – features, problems of different stages of the process of building model.

Work purpose – study the features, consider possible problems and solutions arising in the construction of a machine learning model.

Tasks:

1. an overview of the big data problems and solutions;
2. an overview of the main stages of building machine learning models
3. implementation of the application to display the detected steps and their settings, which allows to obtain a trained model of machine learning and evaluate its accuracy;
4. analysis of the functioning of the application on sets of real data

Research methods – analytical comparison.

Results:

1. the analytical review of the stages of building machine learning models is prepared;
2. the most popular methods within the selected stages are considered;
3. found the advantages and disadvantages as well as the main limitations under which the considered methods can be used;
4. the application is implemented, and its work is demonstrated by the example of the real dataset.

Scope – machine learning, big data.