

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра информационных систем управления

Аннотация к дипломной работе
**«ВЛИЯНИЕ ПРЕДОБРАБОТКИ ДАННЫХ НА РЕЗУЛЬТАТЫ ЗАДАЧИ
КЛАССИФИКАЦИИ»**

Щемелёв Виталий Андреевич

Научный руководитель – старший преподаватель кафедры ИСУ ФПМИ
Мацкевич В. В.

Минск, 2025

АННОТАЦИЯ

Дипломная работа, 42 страницы, 2 рисунка, 1 таблица, 4 приложения, 15 источников.

Ключевые слова: ПРЕДОБРАБОТКА ДАННЫХ, НОРМАЛИЗАЦИЯ, КЛАССИФИКАЦИЯ, НЕЙРОННАЯ СЕТЬ, PYTHON, SKLEARN, ТРАНСФОРМАЦИЯ, ОЧИСТКА ДАННЫХ, DATA SCIENCE

Объект исследования – процессы подготовки и преобразования данных, применяемые перед решением задачи классификации в рамках систем машинного обучения.

Предмет исследования – методы предварительной обработки данных (очистка, нормализация, трансформация, редукция и т.д.) и их влияние на точность и устойчивость моделей классификации, включая нейросетевые подходы.

Цель исследования – выявление и обоснование степени влияния различных подходов к предобработке данных на результаты решения задачи классификации с применением современных алгоритмов машинного обучения.

Методы исследования – а) теоретические – анализ научной и технической литературы по машинному обучению, классификации и инженерии данных; обзор существующих методик предобработки; формализация этапов подготовки данных; б) практические – реализация программного комплекса на языке Python с использованием библиотек scikit-learn и Keras; проведение серии экспериментов с использованием различных методов нормализации и трансформации; обучение нейронной сети на наборах данных с различными вариантами предобработки; сравнение метрик качества классификации.

Полученные результаты – проведён сравнительный анализ эффективности нескольких методов предобработки данных в задаче классификации, выявлено значительное влияние нормализации и масштабирования на точность нейросетевой модели, реализован программный комплекс, позволяющий варьировать методы обработки данных и визуализировать полученные результаты.

Область возможного практического применения – разработка интеллектуальных систем анализа данных, систем поддержки принятия решений, автоматизированных диагностических и экспертных систем, образовательные и исследовательские платформы в области машинного обучения и искусственного интеллекта.

АНАТАЦЫЯ

Дыпломная праца, 42 старонкі, 2 ілюстрацыі, 1 табліца, 4 дадатка, 15 крыніц.

Ключавыя словы: ПАПЯРЭДНЯЯ АПРАЦОЎКА ДАДЗЕННЫХ, НАРМАЛІЗАЦЫЯ, КЛАСІФІКАЦЫЯ, НЕЙРАСЕТКА, PYTHON, SKLEARN, ТРАНСФАРМАЦЫЯ, АЧЫСТКА ДАДЗЕННЫХ, DATA SCIENCE

Аб'ект даследавання – працэсы падрыхтоўкі і пераўтварэння даных, што выкарыстоўваюцца перад вырашэннем задачы класіфікацыі ў сістэмах машыннага навучання.

Прадмет даследавання – метады папярэдняй апрацоўкі дадзеных (ачыстка, нармалізацыя, трансфармацыя, рэдукцыя і інш.) і іх уплыў на дакладнасць і ўстойлівасць мадэляў класіфікацыі, у тым ліку нейрасеткавых.

Мэта даследавання – выяўленне і абгрунтаванне ступені ўплыву розных падыходаў да апрацоўкі дадзеных на вынікі рашэння задачы класіфікацыі з выкарыстаннем сучасных алгарытмаў машыннага навучання.

Метады даследавання: а) тэарэтычныя – аналіз навуковай і тэхнічнай літаратуры па машынным навучанні, класіфікацыі і інжынерыі даных; агляд існуючых метадаў папярэдняй апрацоўкі; фармалізацыя этапаў падрыхтоўкі даных;

б) практычныя – распрацоўка праграмнага комплексу на мове Python з выкарыстаннем бібліятэк scikit-learn і Keras; правядзенне эксперыментаў з рознымі метадамі нармалізацыі і трансфармацыі; навучанне нейрасеткі на датасэтах з рознай ступенню апрацоўкі; параўнанне метрык якасці класіфікацыі.

Атрыманыя вынікі – праведзены параўнальны аналіз эфектыўнасці некалькіх метадаў папярэдняй апрацоўкі дадзеных, выяўлены істотны ўплыў нармалізацыі і маштабавання на дакладнасць нейрасеткі, распрацаваны праграмны комплекс, што дазваляе эксперыментавать з метадамі апрацоўкі і візуалізаваць вынікі.

Вобласць магчымага практычнага прымянення – распрацоўка інтэлектуальных сістэм аналізу дадзеных, сістэм падтрымкі прыняцця рашэнняў, аўтаматызаваных дыягнастычных і экспертных сістэм, адукацыйныя і даследчыя платформы ў сферы штучнага інтэлекту.

ANNOTATION

Diploma work, 42 pages, 2 figures, 1 table, 4 appendix, 15 references

Keywords: DATA PREPROCESSING, NORMALIZATION, CLASSIFICATION, NEURAL NETWORK, PYTHON, SKLEARN, TRANSFORMATION, DATA CLEANING, DATA SCIENCE

Object of the research – the processes of preparing and transforming data used before solving classification problems in machine learning systems.

Subject of the research – methods of data preprocessing (cleaning, normalization, transformation, reduction, etc.) and their influence on the accuracy and stability of classification models, including neural network-based approaches.

Research objective – to identify and justify the impact of different preprocessing techniques on the outcomes of classification tasks using modern machine learning algorithms.

Research methods: a) theoretical – analysis of scientific and technical literature on machine learning, classification, and data engineering; review of existing preprocessing techniques; formalization of data preparation stages; b) practical – development of a software package in Python using scikit-learn and Keras libraries; conducting experiments with different normalization and transformation methods; training a neural network on datasets with varied preprocessing; comparing classification quality metrics.

Results obtained – a comparative analysis of several data preprocessing methods was conducted, significant impact of normalization and scaling on neural network accuracy was identified, and a software tool was developed allowing for flexible preprocessing experimentation and results visualization.

Field of potential application – development of intelligent data analysis systems, decision support systems, automated diagnostic and expert systems, educational and research platforms in artificial intelligence and machine learning.