

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Факультет радиофизики и компьютерных технологий
Кафедра интеллектуальных систем

Аннотация к дипломной работе

**Анализ значимости входных признаков моделей
машинного обучения с учителем**

Раковец Андрей Владимирович

Научный руководитель: старший преподаватель, Щетько Н.Н.

Минск, 2022

РЕФЕРАТ

Дипломная работа: 55 страниц, 2 таблицы, 38 использованных источников.

ОЦЕНКА ВАЖНОСТИ ВХОДНЫХ ПРИЗНАКОВ, ОТБОР ПРИЗНАКОВ, ДЕРЕВО РЕШЕНИЙ, СЛУЧАЙНЫЙ ЛЕС, МЕТОДЫ-ФИЛЬТРЫ, ОБЁРТОЧНЫЕ МЕТОДЫ, ВСТРОЕННЫЕ МЕТОДЫ.

Цель работы: исследование возможности интерпретации поведения обученных моделей машинного обучения с учителем на основе методов оценки важности входных признаков.

В качестве задач исследования, вытекающих из цели, ставилось рассмотреть проблему "черного ящика" при построении моделей машинного обучения с учителем, рассмотреть подход к анализу значимости признаков в моделях на основе деревьев принятия решений, рассмотреть задачу определения значимости признаков в контексте понижения размерности, исследовать алгоритмы на устойчивость к шумовым признакам с различными видами распределений.

Методы исследования: аналитический, статистический.

Новизна заключается в модификации одного из методов оценки важности признаков. Проведенные исследования показали, что данный модифицированный метод склонен увеличивать оценки важности признаков, которые были признаны маловажными оригинальным методом.

РЭФЕРАТ

Дыпломная праца: 55 старонак, 2 табліцы, 38 выкарыстанных крыніц.

АЦЭНКА ВАЖНАСЦІ ЎВАХОДНЫХ ПРЫКМЕТ, АДБОР ПРЫКМЕТ, ДРЭВА РАШЭННЯЎ, ВЫПАДКОВЫ ЛЕС, МЕТАДЫ-ФІЛЬТРЫ, АБГОРТАЧНЫЯ МЕТАДЫ, УБУДАВАНЫЯ МЕТАДЫ.

Мэта даследавання: вывучэнне магчымасці інтэрпрэтацыі паводзін навучаных мадэляў машыннага навучання з настаўнікамі на аснове метадаў ацэнкі важнасці ўваходных прыкмет.

У якасці задач даследавання, якія вынікаюць з мэты, ставілася разгледзець праблему "чорнай скрыні" пры пабудове мадэляў машыннага навучання з настаўнікамі, разгледзець падыход да аналізу значнасці прыкмет у мадэлях на аснове дрэў прыняцця рашэнняў, разгледзець задачу вызначэння значнасці прыкмет у кантэксце паніжэння памернасці, даследаваць алгарытмы на ўстойлівасць да шумавых прыкметах з рознымі відамі размеркаванняў.

Метады даследавання: аналітычны, статыстычны.

Навізна заключаецца ў мадыфікацыі аднаго з метадаў ацэнкі важнасці прыкмет. Праведзеныя даследаванні паказалі, што дадзеныя мадыфікаваны метады схільны павялічваць ацэнкі важнасці прыкмет, якія былі прызнаныя малаважнымі арыгінальнымі метадамі.

ABSTRACT

Thesis: 55 pages, 2 tables, 38 sources.

EVALUATION OF THE IMPORTANCE OF INPUT FEATURES, FEATURE SELECTION, DECISION TREE, RANDOM FOREST, FILTER METHODS, WRAPPER METHODS, EMBEDDED METHODS.

The aim of the research is to investigate the possibility of interpreting the behavior of trained machine learning models with a teacher based on methods for assessing the importance of input features.

As research tasks arising from the goal, it was set to consider the problem of a "black box" when building machine learning models with a teacher, to consider an approach to analyzing the significance of features in models based on decision trees, to consider the task of determining the significance of features in the context of dimension reduction, to investigate algorithms for resistance to noise features with different types of distributions.

Research methods: analytical, statistical.

The novelty lies in the modification of one of the methods for assessing the importance of features. The conducted studies have shown that this modified method tends to increase the estimates of the importance of features that were considered unimportant by the original method.