

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

Аннотация к дипломной работе

ОБНАРУЖЕНИЕ МОШЕННИЧЕСТВА С ПОМОЩЬЮ
МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Лопачук Михаил Юрьевич

Научный руководитель:
ст. преподаватель
К. Г. Атрохов

2016

Дипломная работа содержит 63 страницы, 23 рисунка, 7 таблиц, 33 источника, одно приложение.

АНАЛИЗ ДАННЫХ, БУСТИНГ, ЗАДАЧА КЛАССИФИКАЦИИ, КОМПОЗИЦИОННЫЕ МОДЕЛИ, МЕТРИКИ КАЧЕСТВА, НЕЙРОННЫЕ СЕТИ, НЕСБАЛАНСИРОВАННОСТЬ КЛАССОВ, ОБНАРУЖЕНИЕ МОШЕННИЧЕСТВА, ОПТИМИЗАЦИЯ ГИПЕРПАРАМЕТРОВ.

Объект исследования — алгоритмы машинного обучения, предмет исследования — обнаружение мошенничества с помощью методов машинного обучения. Цель работы: создание программы, способной обнаруживать мошенничество при заказе рекламы в сети Интернет, руководствуясь данными, получаемыми рекламной сетью в момент заказа рекламы.

Особенностью работы является использование композиционных алгоритмов, в том числе градиентного бустинга над решающими деревьями.

Полученные результаты:

- выбрана подходящая метрика качества для задач с несбалансированными классами и вероятностной оценкой;
- построена модель, площадь под ROC-кривой которой 0.99;
- на основе построенной модели создана программа, оценивающая вероятность того, что заказ является мошенническим;
- большое количество различных моделей было обучено и сохранено, они могут быть использованы в аналогичных задачах без настройки.

Дипломная работа носит практический характер.

Дипломная работа выполнена автором самостоятельно.

This thesis project is presented in the form of an explanatory note of 63 pages, 23 figures, 7 tables, 33 references, one appendix.

DATA ANALYSIS, BOOSTING, CLASSIFICATION TASK, ENSEMBLE ALGORITHMS, QUALITY MEASURES, NEURAL NETWORKS, UNBALANCED LEARNING SET, FRAUD DETECTION, MODEL HYPERPARAMETERS TUNING.

In the work various machine learning algorithms are under consideration. The main goal of the thesis is creation of advanced fraud detection system based on the data gathered during client-network transaction process. The main feature of the project is usage of ensemble methods such as gradient boosted trees.

The results are:

- different classification metrics have been discussed, the one that fits best for unbalanced classification tasks has been chosen;
- the model with ROC AUC score 0.99 was constructed;
- based on this model the program with capabilities of object class probability approximation was implemented;
- various models were fitted and saved for later use in other classification tasks of the same field.

The work considers practical aspect of the task.

This thesis project was done solely by the author.