

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра дифференциальных уравнений и системного анализа

Аннотация к дипломной работе

**ОПРЕДЕЛЕНИЕ СТЕПЕНИ БЛАГОНАДЕЖНОСТИ КЛИЕНТА
МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ**

Яблонская Анна Олеговна

Научный руководитель:
кандидат физ.-мат. наук,
доцент А.Э.Малевич

2023

В дипломной работе 39 страниц, 10 рисунков, 3 таблицы, 12 источников.

АНАЛИЗ ДАННЫХ, МАШИННОЕ ОБУЧЕНИЕ, КЛАССИФИКАЦИЯ, РЕШАЮЩИЕ ДЕРЕВЬЯ, СЛУЧАЙНЫЙ ЛЕС, МЕТРИКА, КЛАСТЕРИЗАЦИЯ, АЛГОРИТМ ЛЕЙДЕНА.

Объектом исследования дипломной работы являются клиенты, различающиеся степенью своей благонадёжности.

Основной целью данной работы является автоматизация определения степени благонадёжности клиента.

Для реализации работы использовался язык программирования Python и его библиотеки такие как, Pandas для обработки данных, numpy для дополнительных вычислений, Scikit-learn для построения моделей машинного обучения и вычисления метрик, Matplotlib и Seaborn для построения графиков и визуализации.

В ходе работы была подробно исследована поставленная задача и способы её решения. Был произведён подробный анализ имеющихся данных. Построены три модели машинного обучения, предсказывающие степень риска клиентов. Применённый алгоритм Случайный лес был подробно теоретически разобран, и также описаны его гиперпараметры. Точность каждой модели была посчитана с применением метрики F1-score. Помимо того, в работе описано, почему была выбрана эта метрика. Также для исходных данных были применены методы кластеризации. При этом были рассмотрены, как обычные, так и графовые методы кластеризации.

Новизна работы заключается в применении графовых методом кластеризации для выявления дополнительных связей между клиентами.

Дипломная работа носит исследовательский характер. Ее результаты могут быть использованы в дальнейшем при определении степени благонадёжности клиентов различных организаций при сотрудничестве. (Результаты дипломной работы рекомендованы к внедрению, опубликованию).

Все результаты дипломной работы были обоснованы теоретически. Приведены графики, рисунки и таблицы подтверждающие результаты.

Дипломная работа выполнена автором самостоятельно, все поставленные задачи полностью решены, есть возможность дальнейшего развития исследования.

Дипломная работа выполнена автором самостоятельно.

Thesis project is presented in the form of an explanatory note of 39 pages, 10 figures, 3 tables, 12 references.

DATA ANALYSIS, MACHINE LEARNING, CLASSIFICATION, DECISION TREES, RANDOM FOREST, METRIC, CLUSTERING, LEIDEN ALGORITHM.

The object of study of the thesis work are clients who differ in the degree of their reliability.

The main purpose of this work is to automate the determination of the degree of client's reliability.

To implement the work, the Python programming language and its libraries were used, such as Pandas for data processing, numpy for additional calculations, Scikit-learn for building machine learning models and calculating metrics, Matplotlib and Seaborn for plotting and visualization. The main results of the thesis project are as follows:

In the course of the work, the problem posed and ways to solve it were studied in detail. A detailed analysis of the available data was carried out. Three machine learning models were built to predict the risk level of clients. The applied Random Forest algorithm was theoretically analyzed in detail and its hyperparameters were also described. Next, the accuracy of each model was calculated using the F1-score metric, in addition, the paper describes why this metric was chosen. Also, clustering methods were applied to the initial data. In the work, both conventional and graph clustering methods were investigated.

The novelty of the thesis work lies in the application of graph clustering to identify additional relationships between clients.

This thesis project is a research one. Its results can be used in the future when determining the degree of reliability of clients of various organizations while cooperating. (The results of the thesis are recommended for implementation and published).

All the results of the thesis were justified theoretically. Graphs, figures and tables confirming the results are given.

The thesis project is complete, all tasks have been successfully done, there is a possibility for further research and development.

The thesis work was done by the author independently