

**Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО
ITMO University**

**АННОТАЦИЯ
ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ
SUMMARY OF A GRADUATION THESIS**

Обучающийся / Student Колесникова Мария Алексеевна

Факультет/институт/кластер/ Faculty/Institute/Cluster факультет прикладной информатики

Группа/Group K34212

Направление подготовки/ Subject area 11.03.02 Инфокоммуникационные технологии и системы связи

Образовательная программа / Educational program Программирование в инфокоммуникационных системах 2021

Язык реализации ОП / Language of the educational program Русский

Квалификация/ Degree level Бакалавр

Тема ВКР/ Thesis topic Разработка классификатора компаний на основе анализа ассоциированных открытых проектов

Руководитель ВКР/ Thesis supervisor Капитонов Александр Александрович, доцент, кандидат технических наук, Университет ИТМО, факультет прикладной информатики, доцент (квалификационная категория "ординарный доцент")

**ХАРАКТЕРИСТИКА ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ
DESCRIPTION OF THE GRADUATION THESIS**

Цель исследования / Research goal

Разработка классификатора компаний на основе анализа ассоциированных открытых проектов

Задачи, решаемые в ВКР / Research tasks

1) Разработка программного решения для сбора данных с применением Github API; 2) Подготовка данных (очистка, нормализация, заполнение пропусков); 3) Выбор алгоритма машинного обучения и построение модели классификации; 4) Разработка классификационной модели с использованием Python и библиотек для анализа данных (pandas, scikit-learn и др.); 5) Тестирование и оценка эффективности модели; 6) Анализ полученных результатов; 7) Разработка рекомендаций по улучшению модели.

Краткая характеристика полученных результатов / Short summary of results/findings

Был реализован программный модуль для автоматизированного сбора данных об открытых проектах компаний с платформы Github с помощью Github API. В результате предварительной обработки было собрано и проанализировано 415 компаний и более 15 000 репозиторий, ассоциированных с ними. Проведена нормализация данных, а также реализована инженерия признаков, которые отражают активность, полноту и структурную насыщенность разработки проектов. Проведен анализ данных: исследование распределений, корреляций, выбросов, кластеризация методом KMeans с последующей

интерпретацией групп компаний. На основе подготовленных признаков были обучены несколько моделей машинного обучения (логистическая регрессия, случайный лес, градиентный бустинг и ансамблевый классификатор), был проведен сравнительный анализ. Разработанная система классификации может быть использована для предварительной оценки технологического потенциала компаний, а подход может быть адаптирован для других задач в области анализа открытых данных.

Обучающийся/Student

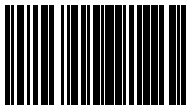
Документ подписан	
Колесникова Мария Алексеевна	
19.05.2025	

(эл. подпись/ signature)

Колесникова
Мария
Алексеевна

(Фамилия И.О./ name
and surname)

Руководитель ВКР/
Thesis supervisor

Документ подписан	
Капитонов Александр Александрович	
19.05.2025	

(эл. подпись/ signature)

Капитонов
Александр
Александрович

(Фамилия И.О./ name
and surname)