



# КАСТЕРИН ИВАН

## АНАЛИТИК ДАННЫХ (СТАЖЕР)

+7 965 468 82 05

kasterin03@bk.ru

г. Москва

### ЦЕЛЬ

Применение навыков в обработке данных, их визуализации и анализа для оптимизации процессов и повышения эффективности компании

### ОБРАЗОВАНИЕ

ФИНАНСОВЫЙ УНИВЕРСИТЕТ  
ПРИ ПРАВИТЕЛЬСТВЕ РФ

2021-2025

ПРЕЗИДЕНТСКИЙ ФМЛ №239

2018-2021

В Г. САНКТ-ПЕТЕРБУРГЕ

ФАКУЛЬТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ  
И АНАЛИЗА БОЛЬШИХ ДАННЫХ,  
"ПРИКЛАДНАЯ МАТЕМАТИКА И ИНФОРМАТИКА"

Образовательная программа "Анализ данных и принятие решений в экономике и финансах"

### НАВЫКИ

- Язык программирования **Python** (анализ данных, машинное обучение, автоматизация задач), основы **Java**
- Инструменты: Jupyter Notebook (Anaconda3), Visual Studio Code, Google Colab, PyCharm
- Библиотеки: NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, SQLite3, OS, Shutil и другие.
- Oracle SQL, PostgreSQL** + проектирование и оптимизация реляционных баз данных
- Power BI** (визуализация результатов анализа, интерактивные дашборды)
- Программы пакета **MS Office**
- Анализ данных и машинное обучение
- Английский язык (B2)

### СИЛЬНЫЕ СТОРОНЫ

- Эффективно планирую** рабочее время, расставляю приоритеты и соблюдаю дедлайны
- Придерживаюсь профессиональной этики и **доброжелательного** подхода в общении
- Умею **находить компромиссы** и решать конфликтные ситуации, поддерживая конструктивный диалог
- Внимателен** к деталям, поддерживаю порядок в документации и процессах
- Нацелен на **качественное** выполнение задач, беру ответственность за результаты

### ПРОЕКТЫ И ДЕЯТЕЛЬНОСТЬ

#### Машинное обучение в задачах акустической классификации механических неисправностей для беспилотных автомобилей

- Создание и очистка датасета звуковых файлов для классификации (Python, Librosa, NumPy, Pandas).
- Обучение и сравнение моделей ML: Decision Trees, Random Forest, SVM, Logistic Regression, Naive Bayes, MLP, k-NN (Scikit-learn).
- Визуализация результатов: метрики точности и матрицы ошибок.

#### Методы интерпретируемой кластеризации эмбедингов сложных объектов

- Подготовка и обработка данных для эмбедингов текстов, изображений, аудио и графов (Python, NumPy, Pandas).
- Генерация эмбедингов с использованием методов Node2Vec и других инструментов (Scikit-learn, NetworkX).
- Применение и сравнение методов кластеризации, оценка качества с использованием метрик (silhouette score, ARI).
- Визуализация кластеров с помощью Matplotlib и других библиотек.
- Определены подходы, обеспечивающие лучшую интерпретируемость и качество кластеризации для различных типов данных.

За время обучения выполнил много практических работ по машинному обучению, работе с данными и построению алгоритмов на языке Python

- Создание образовательного проекта** для подготовки учеников 9-11 классов к экзаменам. Успешное преподавание и наставничество более 3 лет математики и информатики
- Инициировал и возглавил проекты кейс-чемпионата **Changellenge** 2021, 2024