

Токтоналиев Бекмырза

E-mail: toktonaliev.b@phystech.edu * Номер телефона: + 7 977 782 09 98

Data Scientist

Образование

Московский физико-технический институт

2021 - Наст.Время

Физтех-школа биологической и медицинской физики, Биотехнологии

Phystech@DataScience

Февраль.2023 - Декабрь.2023

Окончил курсы Phystech@DataScience, где научился основам прикладной математической статистики, **машинного и глубинного обучения**. В течение курса выполнял домашние задания, которые являются аналогами готовых Технических заданий. В результате я узнал все основные модели машинного обучения и виды архитектур нейронных, получив итоговую оценку **9/10**.

Опыт работы и Проекты

Старт-Ап Platypus

Июль.2023 - Сентябрь.2023

Реализовал код на Python согласно техническому заданию. Код выполняет чтения **файлов MPT** и разделения их на слайсы и маски с помощью библиотек **SimpleITK, Pytorch**.

В результате **скорость** чтения и обработки MPT файлов **увеличилась в 2** раза по сравнению с готовыми аналогами.

Занимался предобработкой полученных данных для обучения написанной мной нейронной сети **Unet** на Pytorch для сегментации, результат которой **IoU ≈ 0.75**

Проекты и Хакатоны

Ai or Not: Классификация - сгенерировано ли изображение при помощи ИИ или создано человеком. Fine tune 6 различных моделей, включая **ViT, Swin** от HuggingFace. С итоговым результатом Logloss ≈ 0.01 , Accuracy ≈ 0.99

Brain tumor: Классификация опухоли мозга и типа опухоли по данным MPT снимков.

Реализовал и натренировал модель **ResNet** с результатом **F1 = 0.75**

Хакатон BG: Определение **саркоматоидности рака** по данным экспессии 20 тыс генов.

Визуализировал данные с помощью **Umap**, решил проблему **Batch Effect'a**, уменьшил размерность обучающей выборки с помощью **PCA**, классифицировал тип рака с помощью **XgBoost** с результатом **F1 = 0.78**

Навыки

Высшая математика
Python

Математический анализ, линейная алгебра, теория вероятностей
Алгоритмы и структуры данных; Numpy, Matplotlib, Pandas, Os, OpenCV

ML

Scikit-learn: Linear regression, Random forest CatBoost/XgBoost

DL

PyTorch | CNN: ResNet,Unet; RNN: LSTM

Дополнительно

Linux, SQL, Latex

Языки

Русский - свободно, Английский - B2, Кыргызский - родной

Олимпиады

Республиканская олимпиада по Биологии
Phystech.International

1 место 2020 г.

Диплом 3 степени

Областная олимпиада по Биологии

1 место 2020г. 3 место 2021 г.

Районная олимпиада по Биологии

1 место 2019, 2020, 2021 г.

National Biology Olympiad

Финалист 2021 год