

Направление: Большие данные, искусственный интеллект, финансовые технологии и машинное обучение

Тема проекта № 1 Разработка приложения для предсказания последующих транзакций клиента на основе банковской истории.

На хакатонах часто предоставляются кейсы от банков, связанные с предсказаниями последующих трат клиентов. Чаще всего такие задачи ориентированы на сам банк и собирают статистику по всем клиентам. Почему бы не попробовать решить похожую задачу, но для отдельного пользователя? Сможем ли мы по банковской истории предсказать траты на следующую неделю? Месяц? А, может быть, даже года?

Цель проекта: Разработать приложение/бот, анализирующие расходы клиента и способное отвечать на запросы о сумме последующих трат.

Основные задачи:

1. Ознакомиться с существующими решениями, изучить источники, провести анализ ситуации.
2. Познакомиться с данными, предоставленными для решения задачи.
3. Расширить датасет с помощью истории банковских транзакций друзей и родственников.
4. Выбрать метрики, которые смогут дать наиболее точный результат.
5. Построить модель, которая на основе предыдущих транзакций за определенный промежуток времени сможет предсказывать последующие траты.
6. Оформить решение в виде бота или приложения для удобного использования.

Материалы для подготовки:

1. Обзор одного из соревнований от Sberbank:
<https://habr.com/ru/article/318160/>
2. Пример задачи и данных для нее:
<https://github.com/dreddsa5dies/DataScienceContest>
3. Решение похожей задачи по предсказанию пола по транзакциям:
https://gh.mltrainings.ru/presentations/Gorkunov_SberbankDataScienceContestA_2016.pdf

Тема проекта № 2 Предсказание курса криптовалют на основе постов в социальных сетях

В настоящее время достаточно значительное влияние на криптосферу оказывают инфлюенсеры, и зачастую они могут своими публикациями оказывать влияние на курс криптовалют! Один из ярких примеров - это твиты Илона Маска, которые спровоцировали взрывной рост Dogecoin. Но можно ли предсказывать колебания курса криптовалют по постам в Twitter?

Цель проекта: Разработать приложение, способное предсказывать курс криптовалюты

Основные задачи:

1. Ознакомиться с существующими решениями, изучить источники, провести анализ ситуации.
2. Сформировать несколько гипотез о возможных факторах, оказывающих влияние на стоимость криптовалюты (например, предыстория, твиты популярных людей, тексты новостей, цена на видеокарты, цена на электричество, ...).
3. Выбрать криптовалюты, для которых будет проводиться разработка и тестирование.
4. Для выбранных криптовалют найти данные о истории колебаний их курса.
5. Построить модели, учитывающие выбранные факторы. Например, одной из самых простых будет являться модель LSTM для предсказания курса по предыстории (ссылка №2 в материалах). Также, например, можно анализировать тексты твитов на частоту упоминания названия криптовалюты и определять эмоциональную окраску текста твита при помощи методов NLP. Все ограничено лишь вашей фантазией.
6. Оценить качество полученных моделей, сделать выводы о влиянии каждого из факторов на цену.

Материалы для подготовки:

1. Создание программы анализа настроений Twitter
<https://www.machinelearningmastery.ru/creating-the-twitter-sentiment-analysis-program-in-python-with-naive-bayes-classification-672e5589a7ed/>
2. Нейросеть для предсказания цены биткоина с использованием LSTM
<https://proglib.io/p/bitcoin-price>

3. Твиты, классифицированные по эмоциональной окраске

<https://www.kaggle.com/kazanova/sentiment140>

4. История курса биткоина

<https://www.kaggle.com/mczielinski/bitcoin-historical-data>

5. Как лидеры мнений Twitter влияют на курсы криптовалют?

<https://vc.ru/crypto/177806-kak-lidery-mneniy-twitter-vliyayut-na-kursy-kriptoalyut>

Тема проекта № 3 Нейросеть, которая увеличивает FPS

В YouTube существует множество видеороликов следующего вида: вот Нью Йорк 1911 года, было так: плохое качество, много шума и мало цвета, МАГИЯ и стало 4К-разрешение и 50/60 кадров в секунду. Или улица Тверская в 1896 году — ЩЁЛК — и она преобразается!

Всё дело в нейросетях, в основе которых лежит интерполяция кадров.

Мультипликация - очень дорогая вещь, так как труд профессиональных художников высоко ценится. Соответственно, увеличение FPS (кадровой частоты) в два раза увеличивает и затраты на создание мультипликации. В данном кейсе вам предлагается дорисовать кадры мультфильма так, чтобы их происхождение не было заметно.

Цель проекта: построить модель для увеличения плавности анимации мультфильма в 2 раза.

Основные задачи:

1. Ознакомиться с принципом работы алгоритма DAIN, изучить источники.
2. Построить модель для интерполяции промежуточного изображения на основе двух данных.
3. Изучить составляющие алгоритма DAIN и степень их влияния на результат интерполяции.
4. Построить модель для увеличения частоты кадров в видеоролике.
5. Возможно ли модифицировать алгоритм так, чтобы “малой кровью” сократить затраты на вычислительные мощности?

Материалы для подготовки:

1. Обзор алгоритма от [Droider.ru](https://droider.ru).
<https://droider.ru/post/dain-neyroset-kotoraya-velichivaet-fps-27-04-2020/>
2. [Реализация](https://github.com/MortenHannemose) подобного алгоритма от MortenHannemose.
<https://github.com/MortenHannemose/pytorch-vfi-cft>
3. [Реализация](https://github.com/baowenbo) DAIN.
<https://github.com/baowenbo/DAIN>

Тема проекта № 4 Рекомендательная система для построения шахматных тренировок

В этом году состоялся матч за звание чемпиона мира по шахматам между нашим соотечественником Яном Непомнящим и чемпионом мира Магнусом Карлсеном.

Не секрет, что подойти в хорошей форме к матчу требует большой и разносторонней работы, которая требует как знания множества шахматных нюансов, так и физической и психологической подготовки.

У таких шахматистов, как правило, есть большая команда, которая помогает им становиться сильнее каждый день.

Но что делать обычному шахматисту-любителю, который играет, например, на платформе lichess или chess.com? Как ему совершенствоваться?

Цель проекта: Разработайте модель, которая позволит игроку понять, какие стороны у него самые слабые и над чем ему надо работать

Основные задачи:

1. Ознакомиться с существующими решениями, изучить источники, провести анализ ситуации.
2. Познакомиться с материалами, предоставленными для решения задачи.
3. Определите, с помощью каких метрик можно описать силу игрока.
4. Используя данные платформы lichess <https://database.lichess.org/> постройте для заданного игрока «паутинную» диаграмму силы игрока.
5. Выберите метрики, которые влияют на исход партии больше всего.
6. На основе этих данных предложите возможные виды тренировок/книг подходящих для данного игрока.

Материалы для подготовки:

1. Платформа для анализа игры шахматиста: <https://aimchess.com/>
2. Пример проекта а анализ качества игры на базе партий платформы lichess:
https://www.matthewbellanalytics.com/pages/jupyter_html/jupyter_chess.html
3. Данные платформы lichess <https://database.lichess.org/>