

Шуточный Discord-бот с нешуточным содержанием

Общее описание

В основе нашего бота лежит обученная нами модель машинного обучения GradientBoostingClassifier из библиотеки sklearn. Модель прогнозирует направление движения курса доллара на сегодня по прошлым данным.

Для прогноза, взяты следующие признаки:

- исторические данные по температуре воздуха в Москве
- исторические данные по числу заболевших COVID в России
- исторические данные по числу заболевших COVID в Москве
- исторические данные по числу смертей от COVID в России
- исторические данные по числу смертей от COVID в Москве
- исторические данные по курсу доллара

Прогноз, по команде, бот показывает пользователю.

Обучение модели и сбор данных — этапы

- Сбор исторических данных по разным источникам.

Исторические данные для обучения модели должны были соответствовать тем данным, которые мы, затем, сможем в режиме реального времени получить из API для подачи в модель и получения прогноза. Причем, из API мы должны получать прошлые данные для получения прогноза на сегодня.

Для получения данных мы использовали:

API погоды: <https://yandex.ru/dev/weather/#start>

API статистика коронавируса: <https://github.com/M-Media-Group/Covid-19-API>

API курс: <https://www.cbr-xml-daily.ru/>

API фото: [https://api.unsplash.com/photos/random?](https://api.unsplash.com/photos/random?client_id=0WFgkAcPaQN54A4ZUELi1ZfT3LyBZK_8FEMyVm7aLL0&query)

[client_id=0WFgkAcPaQN54A4ZUELi1ZfT3LyBZK_8FEMyVm7aLL0&query](https://api.unsplash.com/photos/random?client_id=0WFgkAcPaQN54A4ZUELi1ZfT3LyBZK_8FEMyVm7aLL0&query)

Данные по ковид для датасета: <https://github.com/CSSEGISandData/COVID-19>

Данные по курсу для датасета:

Погодные данные для датасета: <https://www.visualcrossing.com/weather/weather-data-services>

- Очистка данных и создание таргета

Таргет создавался из двух классов [0, 1].

0 — падение курса

1 — не падение курса

- Разбиение на тренировочную и тестовую выборки

На тренировочную и тестовую выборки данные были разбиты до предобработки данных, чтобы не тестовые и тренировочные данные ни чего «не знали» друг о друге. Таким образом,

результат обучения и теста приближен к реальной работе модели на неизвестных для модели данных.

- Предобработка данных

Предобработка делалась отдельно для тренировочной и тестовой выборок.

Все наши признаки - непрерывные, но находятся в разных шкалах по размеру, например, температура воздуха и количество заболевших по России. Поэтому была сделана стандартизация признаков.

- Подбор гиперпараметров и кросс-валидация для нашего алгоритма ML

Подбор гиперпараметров и кросс-валидация делались с помощью метода GridSearchCV().

- Построение модели и обучение с помощью кросс-валидации

После получения гиперпараметров модель была построена с данными параметрами и обучена, используя кросс-валидацию.

- Сохранение модели

Обученная модель была сохранена.

Бот

Для общения с Discord-ботом присоединитесь к каналу по ссылке:

<https://discord.gg/UYeD93GC>

Команды для бота:

.данные или .прогноз - бот выдает данные для модели машинного обучения на текущую дату (статистика по ковид, погодные данные, курс доллара), а также прогноз по курсу доллара на завтра - упадет или не упадет курс доллара в рублях.

.погода <город> - бот выдает текущую погоду в указанном городе

пример: .погода Санкт-Петербург

.ковид <country> - бот выдает количество заболевших ковидом в указанной стране.

пример: .ковид Canada

.курс <currency> - бот выдает текущий курс указанной валюты в рублях

пример: .курс jpy

.фото <collections> - бот выдает случайно фото на указанную тему

пример: .фото ocean

.перезагрузка – бот снимает сглазы и восстанавливает моральные и физические силы

Общая схема процессов работы бота

