

ML для оценки параметров экономических моделей

Кейсодержатель



Банк России

Команда

swaepnet

Организатор

цифровой
прорыв



сезон: III

Задача хакатона

Предсказание параметров моделей, путем анализа ключевых макроэкономических показателей.

Создание быстрого и точного алгоритма

для анализа макроэкономических показателей.

Оптимизация прогностических моделей

с использованием нейронных сетей.

Разработка вероятностного алгоритма

оценки параметров

НАШ ПОДХОД К РЕШЕНИЮ

01

Анализ задачи и данных, EDA, выбор стратегии по решению задачи.

02

Подготовка и предварительная обработка данных, разработка baseline модели

03

Разработка и обучение нейронной сети на базе CNN и бустинг модели

04

Фич-инжиниринг

05

Стекинг двух моделей. Подбор гиперпараметров

ОСНОВА РЕШЕНИЯ

Нейронная сеть для предсказания вектора значений для каждого параметра

Киллер фичи

Уникальность и проработка

EDA

Проведен глубокий анализ задачи и данных, выявлены закономерности (например сезонность), определены связи в данных и стат. метрики данных

Две модели

Использование двух принципиально различных моделей для решения общей задачи: нейронная сеть CNN + бустинг.

Уникальная модель

Нейронная сеть для каждого искомого значения предсказывает не одно значение а вектор допустимых значений, на основе которого рассчитываются итоговые характеристики

ФичИнжиниринг

Реализованы дополнительные временные ряды к существующим. Сгенерировано более 120 дополнительных признаков.

Для каждого параметра своя модель

Для каждого из 15 параметров обучена собственная модель

1

2

3

4

5



УНИКАЛЬНОСТЬ РЕШЕНИЯ

1

ДВЕ РАЗНЫЕ МОДЕЛИ

Использование двух
принципиально различных моделей
для решения общей задачи.

2

НЕЙРОННАЯ СЕТЬ

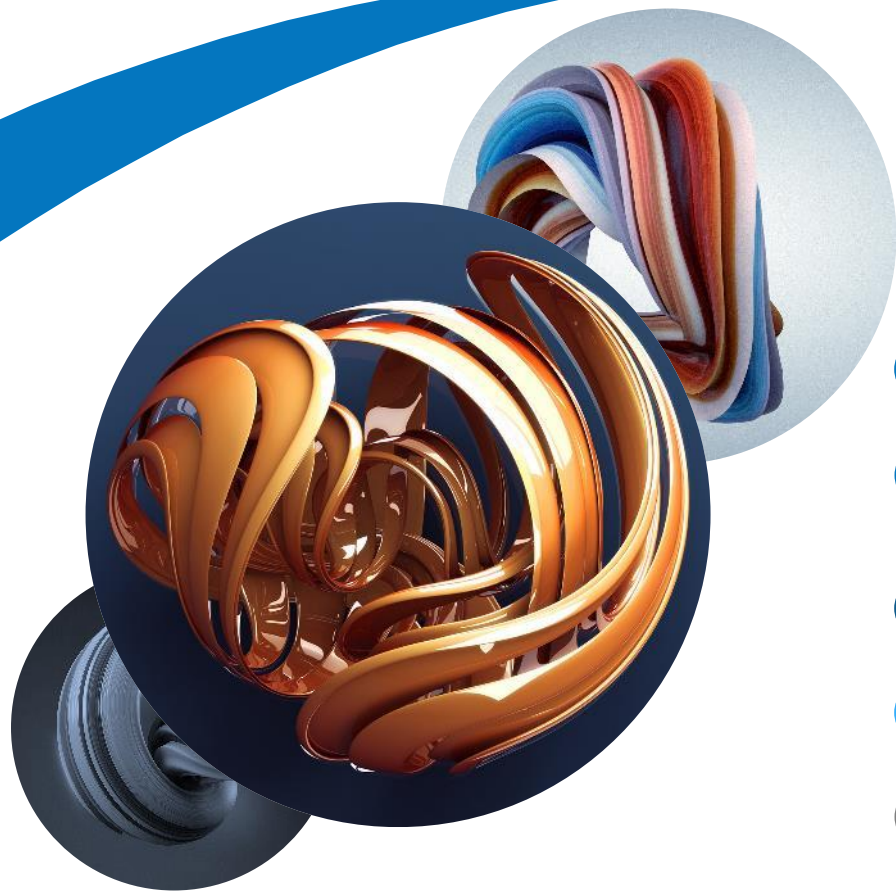
предсказывает допустимый
вектор значений
для расчета итоговых параметров

3

НЕЗАВИСИМЫЕ МОДЕЛИ

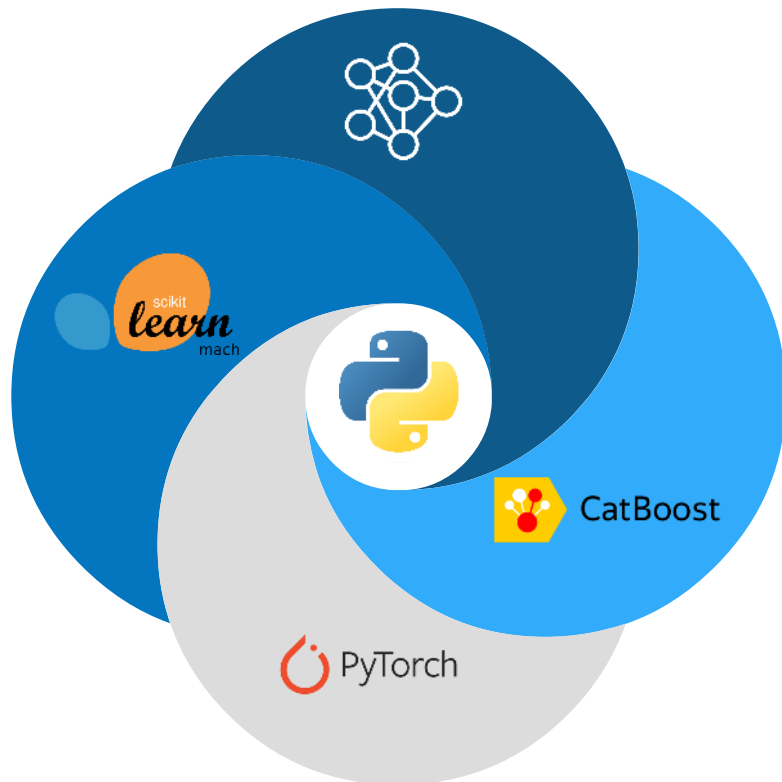
Построены независимые модели
для каждого из 15 признаков

Пробовали, но не пошло



- ✓ Ансамбли и бустинг моделей
- ✓ Применение классических методов (например, байесовский подход)
- ✓ Аугментация и синтетика данных
- ✓ Внедрение Attention! блоков
- ✓ Стэкинг двух моделей

ИНСТРУМЕНТЫ И ФРЭЙМВОРКИ



СТЭКИНГ МОДЕЛЕЙ



СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ



ГРАДИЕНТНЫЙ БУСТИНГ



СТАНДАРТНЫЕ ИНСТРУМЕНТЫ PYTHON

Jupyter, Numpy, Pandas, Matplotlib, Google Colab и многие другие...

Что можно улучшить?



НОВЫЕ АРХИТЕКТУРЫ

Рекуррентные сети, Генеративные модели и пр.



ФИЧИНЖИНИРИНГ/АУГМЕНТАЦИЯ

Генерация агрегированных фичей, создание синтетических временных рядов.



ОБРАТНОЕ ФОРМИРОВАНИЕ РЯДОВ

Из предиктов пытаемся обратно восстановить ряды и проверяем уровень совпадения рядов (GAN метод).



КОМБИНАЦИЯ ВСЕХ МЕТОДОВ

Комплексный подход к решению задачи: глубокий анализ данных, новые фичи, ансамблирование моделей, применение state of the art (SOTA) практик.

Команда Sweepnet

Пожалуй, лучшая команда на любом хакатоне

ДМИТРИЙ
КОЖЕМЯКО
Капитан

 @Dimk_88



ИВАН ЧЕРНЫХ
Data Scientist

 @iceman_o_0



РУСЛАН ЧЕРНЕНКО
Аналитик

 @RChernenko



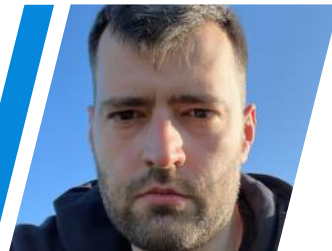
ДАРЬЯ ОРЛОВА
Аналитик

 @darya_or_lo_va



СЕРГЕЙ
БОРИСОВСКИЙ
Разработчик

 @CodeDrivenA



Спасибо!

Решение: <https://github.com/DimkKozhem/The-model-of-the-client-s-propensity-to-purchase-a-parking-space>



по всем вопросам обращайтесь к капитану
команды Sweepnet
Дмитрий Кожемяко

 @Dimk_88



Банк России

цифровой
прорыв 

сезон: III

