**Прогнозирование цен ДДУ по исходным данным**

Программа реализована на Python 3.7.3 и включает в себя 3 основных модуля:

- dnn\_class.py;

- fit\_model2.py;

- forecast.py.

В первом модуле (**dnn\_class.py**) описан класс используемой сети глубокого обучения с помощью модуля Pytorch. Для прогнозирования цен ДДУ используется сеть Ворда со следующей архитектурой:

1. Первый скрытый слой состоит из 3 блоков, каждый из которых имеет 44 нейрона. В каждом блоке используется своя активационная функция: softplus(), logsigmoid(), tanh().
2. Второй скрытый слой состоит из 9 блоков по 22 нейрона каждый. Каждые три блока соединены с одним предыдущим. В каждом блоке используется своя активационная функция: softplus(), logsigmoid(), tanh().
3. Третий скрытый слой состоит из 3 блоков, каждый из которых имеет 11 нейронов. Каждый блок соединен с тремя предыдущими. В каждом блоке используется своя активационная функция: softplus(), logsigmoid(), tanh().
4. Последний слой принимает сигналы с трех предыдущих блоков и входного слоя. К данному слою применяется единственная активационная функция ReLU().

Данная архитектура нейронной сети позволяет ускорить сходимость алгоритма до нужной величины точности и генерировать новые переменные посредствам распараллеливания блоков и применения различных активационных функций в каждом блоке. Применение различных методов обработки для одного и того же набора данных позволяет сказать, что нейронная сеть анализирует данные с различных аспектов.

Во втором модуле (**fit\_model2.py**) реализованы необходимые функции для препроцессинга данных и обучения нейронной сети. Точность прогнозирования измеряется по формуле:

Детальное описание средней абсолютной ошибки MAPE можно найти здесь: <https://en.wikipedia.org/wiki/Mean_absolute_percentage_error>

При измерении средней ошибки используются результаты валидации методом Repeated random sub-sampling validation. Подробности метода можно посмотреть здесь: <https://en.wikipedia.org/wiki/Cross-validation_(statistics)>

Результаты обучения показывают, что модель сходится к средней валидационной точности 95% примерно на 1000-ой итерации.

Модуль (**forecast.py**) предназначен для получения прогнозов по новым данным. Формат представления новых данных предложен в файле input\_example.xlsx.

После прогноза создается файл output.xlsx в котором приводятся исходные данные и прогнозируемая цена по ДДУ (по каждой позиции).

**В настоящее время** модель обучена и достаточно использовать только файл **forecast.py** для получения результатов!

В папке sd хранятся служебные файлы, которые генерируются на этапе обучения модели.

**Возможные проблемы:**

1) Несовпадение имен колонок в обучающем наборе данных и в наборе данных для прогнозирования. Необходимо откорректировать колонки для совпадения наименований.

2) Появления новых значений в категориальных признаках. В этом случае необходимо либо переобучить модель на новых данных, либо пометить новые значения признаков как пропущенные (удалить их из ячейки Excel, сделать ячейку пустой).