Техническое задание по кейсу «Прогнозирование рыночных цен на арматуру»

Предыстория

Одна из основных статей расходов девелоперов – это закупка материалов. При этом, среди общих затрат на материалы около 10% приходится на арматуру.   
Упрощенно, процесс закупки материалов выглядит так:  
Категорийный менеджер Иван, придя в пятницу на работу, устраивает тендер на закупку арматуры объема Х (для простоты давайте считать, что потребность ГК Самолет в арматуре постоянна, нам нужно Х тонн в неделю). Например, он может организовать тендер на объем 5Х, или на объем 2Х.   
Он знает какая цена на арматуру сегодня, но не знает, какая она будет через неделю.

В принятии решение о размере проводимого тендера может помочь прогнозная модель.  
Например, если модель дает прогноз, что через неделю рыночная цена на арматуру снизится – то закупаться стоит на 1 неделю, и наоборот, если модель прогнозирует тренд на рост цены – то следует сделать более масштабную закупку.

С одной стороны, прогнозирование цены – это стандартная модель временных рядов, но бизнесу же важен не просто временной ряд, а качество предлагаемых «бизнес»-решений: рекомендация необходимого объема, поэтому результатом работы модели должно быть целое число N – на сколько недель вперед проводить закупку (1 <=N <=10)

Постановка задачи

Цель: Помочь категорийному менеджеру сэкономить  
В приложении есть понедельная история рыночных цен на арматуру в промежутке 2019-2022 гг.  
Необходимо разработать модель, которая по истории [1,T] делает рекомендацию по объему тендера для недели T. Рекомендация на N недель фиксирует цену на N недель, следующий тендер будет проведен на неделе T+N. Считаем, что тендеры проводятся по пятницам, оперировать можно историческими данными до четверга включительно.

Целевая метрика: средняя цена закупки арматуры при проведении тендеров на промежутке 1/07/22 – 23/12/22

Методология расчета (псевдокод):  
pp\_array = []  
ll\_array = []  
for start\_date in [все пятницы с 1/07 – 23/12]:  
 for end\_date in [start\_date + 2 недели, 26/12]:  
 price\_period = Средняя Цена закупки за период [start\_date, end\_date)  
 len\_period = end\_date – start\_date – 1  
 pp\_array.append(price\_period)  
 ll\_array.append(len\_period)  
final\_avg\_price = sum(np.array(pp\_array) \* np.array(ll\_array)) / sum(np.array(ll\_array)) – взвешенное среднее

Стоит аккуратно отнестись к вычислению средней закупки за период, имейте в виду что в остаток периода может поместиться не весь N.   
На забудьте аккуратно настроить схему валидации, подумайте над отбором признаков, точек мало – данных много.

Данные, которые вам могут понадобиться

1. Данные по ценам:  
   
2. <https://cbr.ru/> - ключевая ставка ЦБ РФ, курс доллара
3. Данные ДомРФ, например:  
   https://xn--80az8a.xn--d1aqf.xn--p1ai/%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D1%82%D0%B8%D0%BA%D0%B0/%D0%B2%D0%B2%D0%BE%D0%B4\_%D0%B6%D0%B8%D0%BB%D1%8C%D1%8F/%D0%B4%D0%B5%D1%82%D0%B0%D0%BB%D0%B8/%D1%82%D0%B0%D0%B1%D0%BB%D0%B8%D1%86%D0%B0
4. Секретные данные от организаторов:
   1. Данные по рыночным показателям  
      
   2. Данные по производству, потреблению и проч.:  
      
5. <https://bi.gks.ru/biportal/contourbi.jsp?allsol=1&solution=Dashboard&project=%2FDashboard%2FPrices>  
   Средние потребительские цены на непродовольственные товары -> таблица -> топливо
6. Все, что найдете в интернете

Фреймворки, которые могут пригодиться при разработке моделей

1. <https://unit8co.github.io/darts/> - решение для работы с временными рядами
2. <https://etna.tinkoff.ru/> - еще одно решение для работы с временными рядами
3. <https://stable-baselines3.readthedocs.io/en/master/> - решение для работы с RL-агентами
4. <https://catboost.ai/> - решение для построения бустинговых алгоритмов
5. <https://tsfresh.readthedocs.io/en/latest/> - решение для генерации доп. признаков для работы с временными рядами

Неочевидные способы, которыми можно попробовать решать

У нас есть среда (Environment) – история рыночных цен + остальные данные  
Непонятно, как от прогноза стандартных моделей (модели временных рядов, каскад регрессий) перейти к рекомендации объема  
Может быть, можно построить RL-агента?  
action – объем закупки, state – данные по состоянию на текущую дату, reward – ср. цена закупки за период.  
Дерзайте!