

Университет Искусственного Интеллекта - декабрь 2020

# Пальмовое масло

Предсказание цены на 30 дней вперед

\*Не нашел картинку с пальмами ))





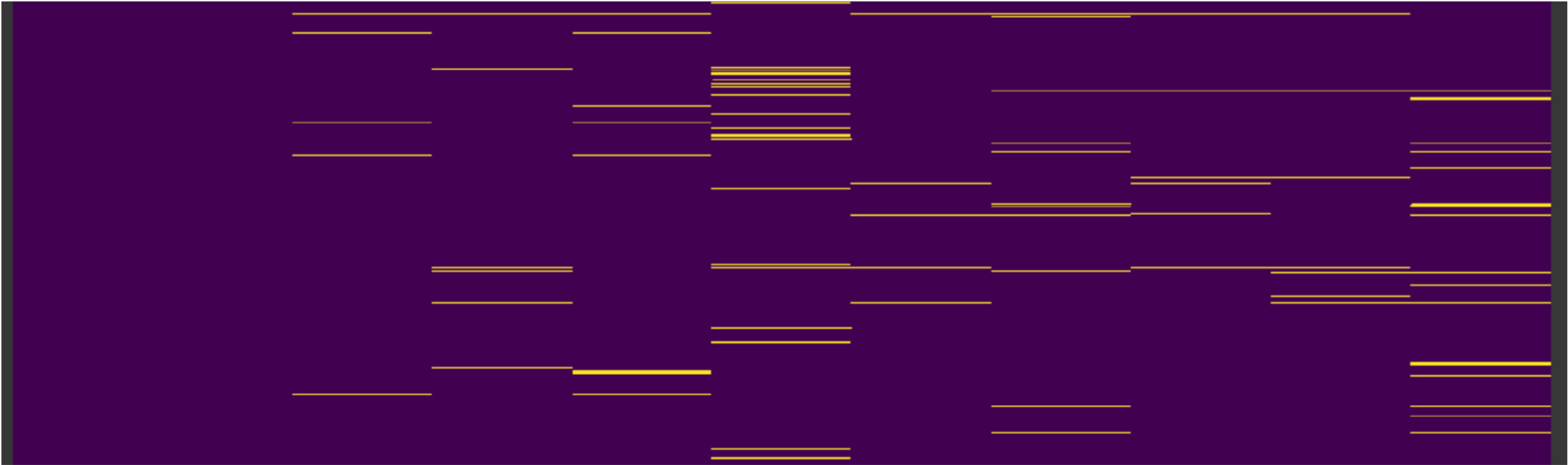
# Формулировка задачи

- Предсказать цену и/или тренд на Пальмовое масло на 30 дней вперед
- Дан датасет с ценой и ценами на сопутствующие товары, с которыми, может быть связь

# Исходный датасет

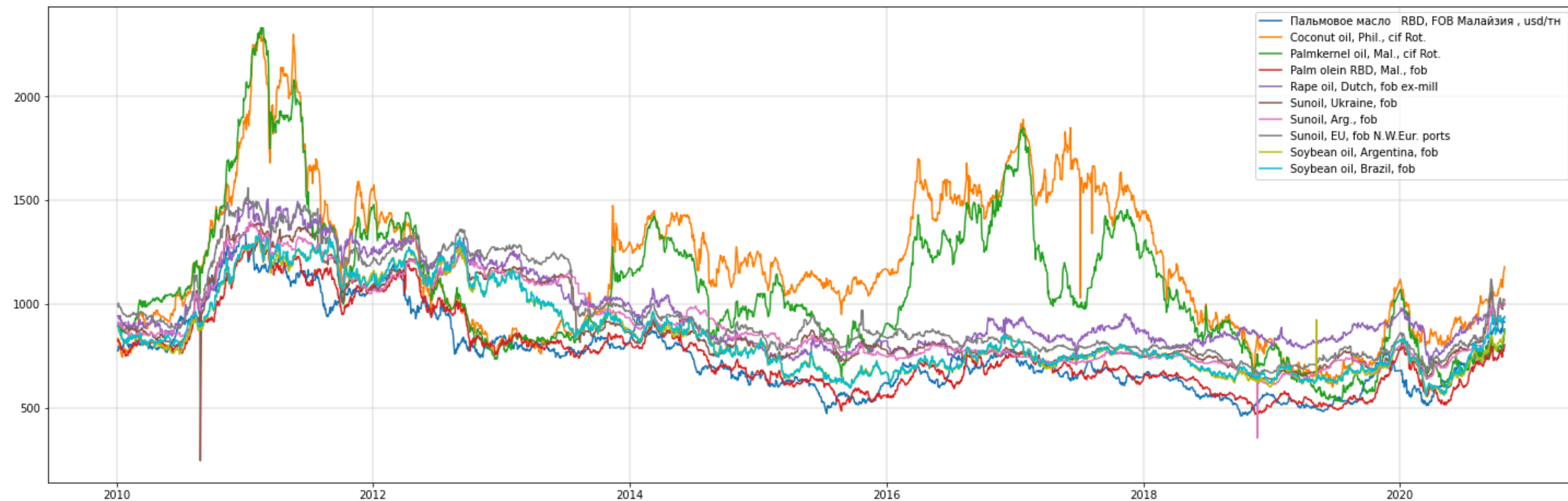
	дата	Soybean oil, Brazil, fob	Soybean oil, Argentina, fob	Sunoil, EU, fob N.W.Eur. ports	Sunoil, Arg., fob	Sunoil, Ukraine, fob	Rape oil, Dutch, fob ex-mill	Palm olein RBD, Mal., fob	Palmkernel oil, Mal., cif Rot.	Coconut oil, Phil., cif Rot.	Пальмовое масло RBD, FOB Малайзия, usd/тн
2	2010-01-04 00:00:00	896	891	990	NaN	NaN	942	822	900	835	795
3	2010-01-05 00:00:00	895	890	1000	910	NaN	945	830	900	830	805
4	2010-01-06 00:00:00	891	887	1005	915	NaN	943	830	900	825	805

# Количество артефактов



# Рабочая идея

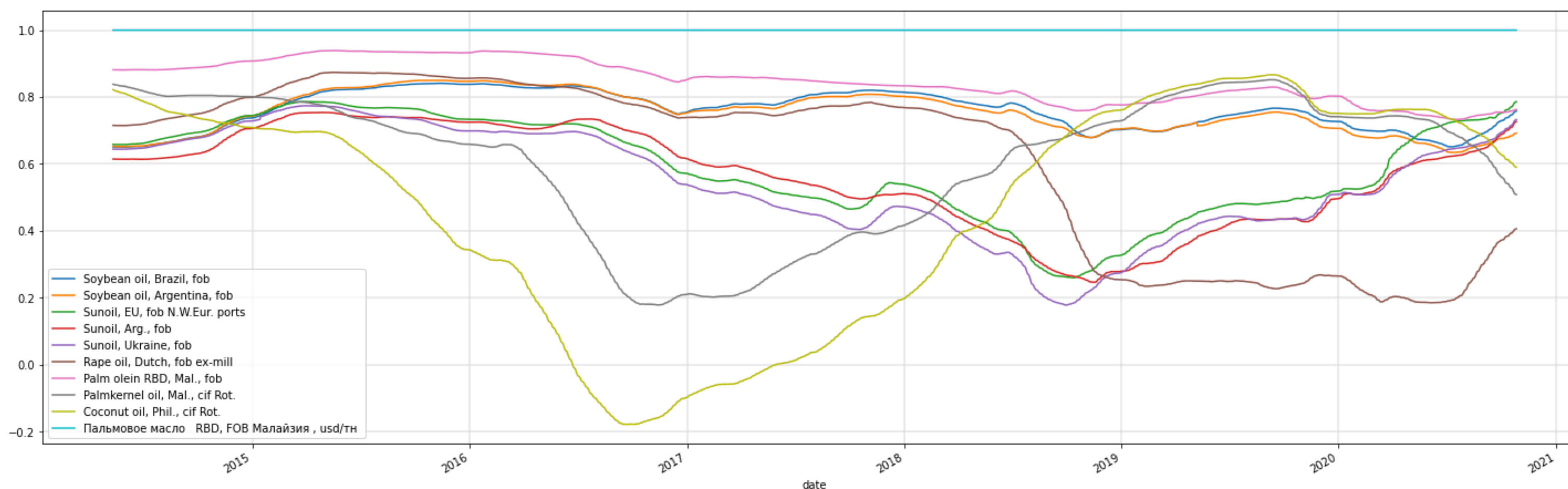
- Предсказать цену на большой промежуток времени достаточно сложно
- Чтобы понять, имеет ли задача решение, мы сдвинули цены на пальмовое масло на 30 дней назад и решили посмотреть, есть ли, из нашего набора факторов те, которые имеют значимую связь с ценой, через 30 дней



# Результаты анализа датасета

## rolling.correlation - 3 летняя плавающая корреляция

- Факторы влияние факторов на  $Y$  не постоянно!
- Но! Есть факторы, имеющие связь около 80% с будущей ценой
- При обучении нейронки будет необходимо это учесть

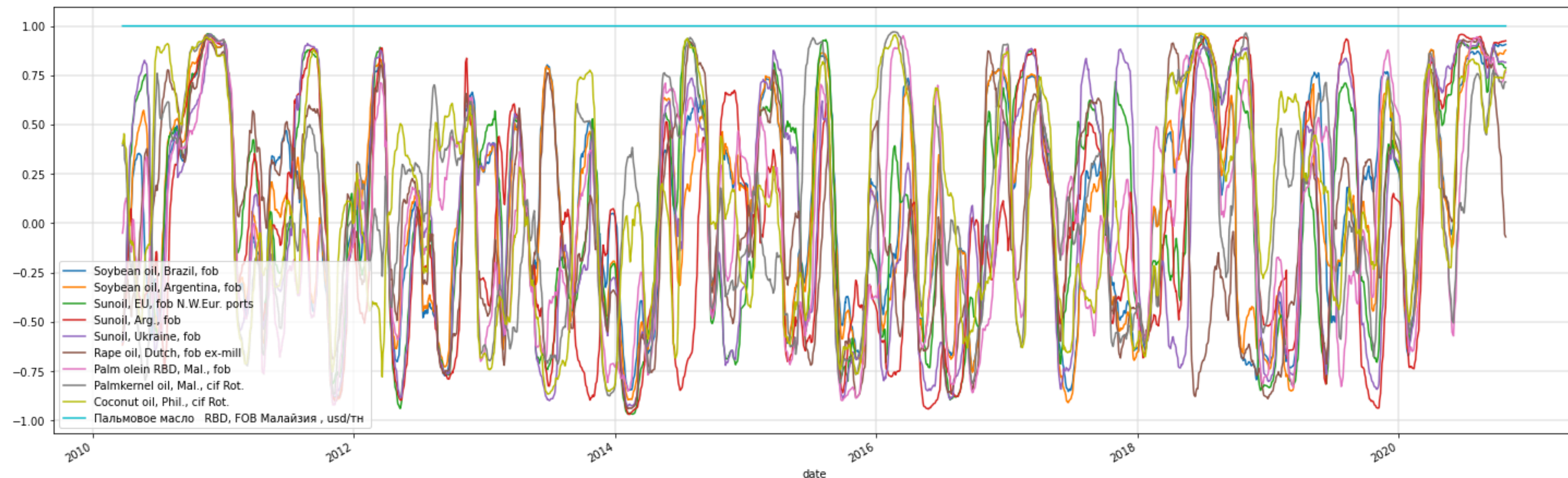




# Результаты анализа датасета

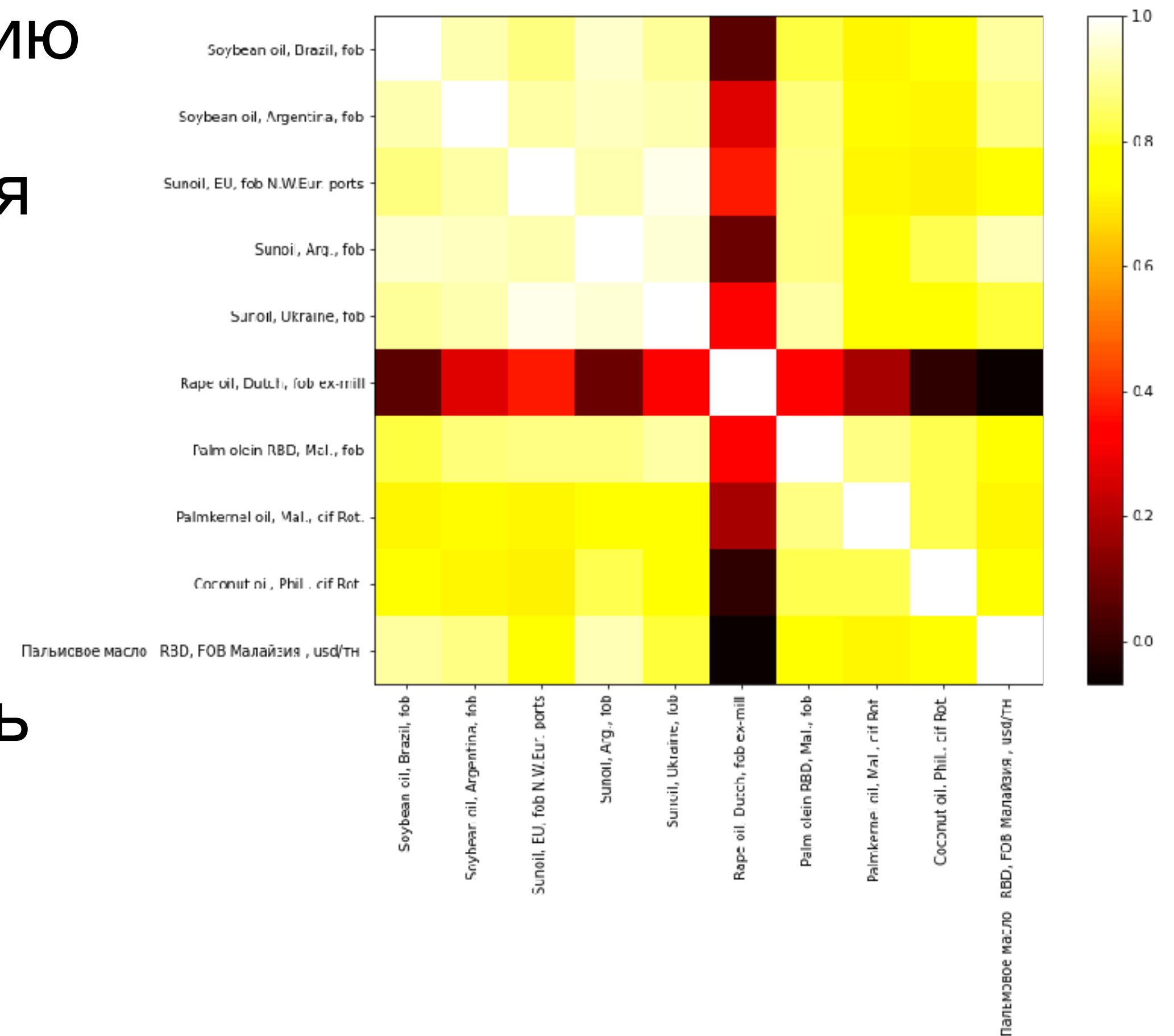
## rolling.correlation - 60 дневная плавающая корреляция

- Факторы регулярно меняют силу и направление своего влияния на Y
- Но! Есть факторы, имеющие связь около 80% с будущей ценой
- При обучении нейронки будет необходимо это учесть



# Принятое решение

- Провести ряд экспериментов по удалению факторов, оказывающих слабое или негативное влияние на процесс обучения нейронной сети.
- Экспериментом определить интервал, который наиболее эффективен для обучения нейронки
- Например: на последних 60 днях удалить нужно только 1 фактор:





# Результат работы

Было создано три ноутбука:

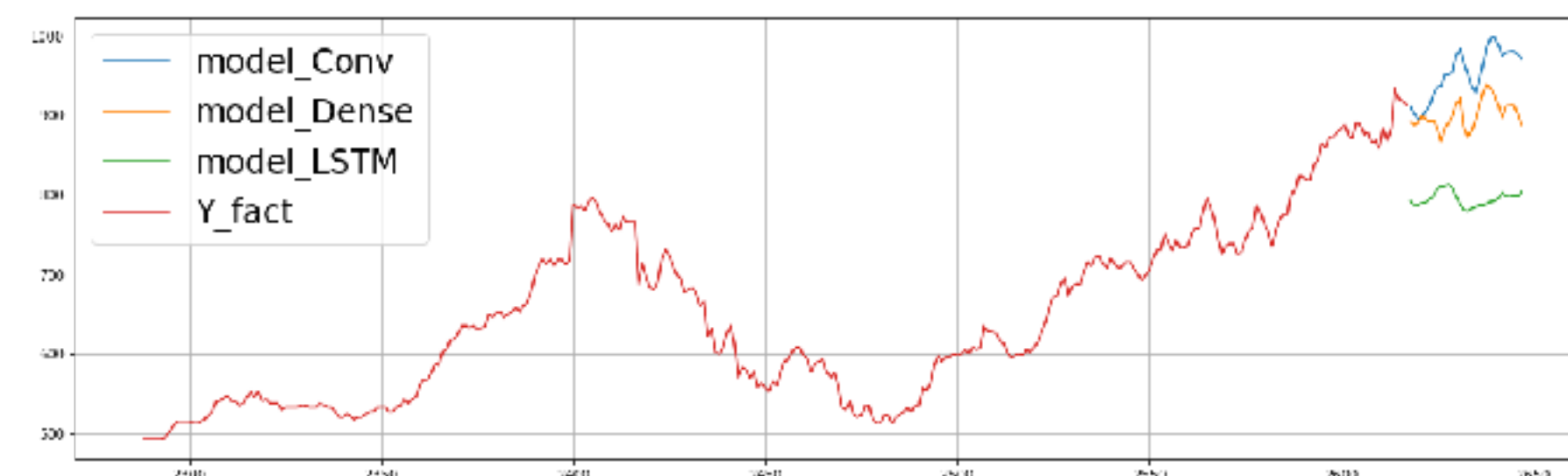
1. Подготовка данных, очистка датасета от артефактов
2. Создание и обучение нейронной сети
3. Предикт

# Результат и направления дальнейших исследований

## Пилотный проект подтвердил решаемость задачи

- Пути улучшения пилота:

1. Подбор архитектуры нейронной сети
2. Подбор ансамблей (пакетов данных)
3. Более глубокий анализ влияния факторов и особенности их учета в прогнозировании цены. К сожалению, сейчас анализ выявил "неприятную" особенность: факторы меняют влияние на цену от сильного положительного влияния до сильного отрицательного в зависимости от периода (см файл подготовки датасета)
4. Проведение форвардного анализа (чтобы понять необходимость и частоту переобучения нейронки)



Предсказываем ТРЕНД на 30 дней вперед:

prediction	accuracy	model_name
131 Up	0.878788	PalmOil_trend_Conv_1h.h5
131 Up	0.810606	PalmOil_trend_Dense_1h.h5
131 Up	0.848485	PalmOil_trend_LSTM_1h.h5