

Университет ИТМО

Проектная работа
"Прогнозирование цен на акции с помощью
рекуррентных нейросетей"
по дисциплине «Визуализация и моделирование»

Команда: Кумпан Виктор Викторович,
Легин Денис Анатольевич

Поток: 1.2

Группа: К3241-40

Факультет: ИКТ

Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

Аннотация

В данной работе начнем с выбора одного датасета из тех, с которыми работали участники команды и для которого применима задача глубокого обучения с учителем. Далее проанализируем данные по CRISP-DM, а также оценим модель.

1 Выбор датасета

Перед нами выбор одного из двух датасетов, а именно "Статистика суицидов ВОЗ" и "Акции американских компаний". Наш выбор был сделан в сторону второго датасета по следующим причинам:

- Формат данных идеально подходит для задачи регрессии
- Достаточный объем данных для задачи регрессии
- Стабильность данных
- Возможно выделить определенные зависимости => методы глубокого обучения смогут в какой-то мере найти их и определенные паттерны в них.
- Возможность определить последующую оценку качества моделирования и прогнозирования данных.

2 Анализ данных по CRISP-DM

Проанализируем выбранные нами данные "Акции американских компаний" по следующим 6 фазам:

1. Бизнес-анализ.

I Цели проекта Основные пользователи продукта это частные лица, либо компании (B2B, B2C). Бизнес-цель проекта - автоматизированная поддержка принятия решений по операциям с акциями различных компаний. На данный момент существуют компании по прогнозированию стоимости акций к примеру [StocksNeural](#), по данным с их сайта точность прогноза составляет порядка 90% и они переигрывают индекс S&P500 ([IKnowFirst](#) и [StockCharts](#))

II Текущая ситуация На данный момент мы имеем облачный сервис для вычислений "Google Colab" в котором будем производить предварительное обучение модели. Данный будет храниться на локальном компьютере MacBook Pro. Вероятные риски проекта: не получить модель, которая переигрывает хотябы любой индекс и сможет принести определенную прибыль.

III Решаемые задачи с точки зрения аналитики. Основной метрикой для оценки качества модели служит RMSE и R^2 . Минимальный порог $R^2 = 0.85$ и $RMSE = 1.5$, оптимальный $R^2 = 0.92$ и $RMSE = 0.8$

2. **Анализ и подготовка данных** Данные раздел упомянуты в следующих статьях: [1 статья](#) [2 статья](#). Возьмем только данные по следующим компаниям из различных секторов Apple, Nike, ExxonMobil.

3. Моделирование

I Выбор алгоритма. Вернемся к решаемой задачи - предсказание временных рядов (задача регрессии), основным инструментом в глубоком обучении для данной задачи являются рекуррентные нейронные сети с долго краткосрочной памятью. Отличной архитектурой являются LSTM [1](#), с ними и будем в дальнейшем работать. Данные, которые мы будем подавать на вход - числовые ряды (стоимости акций на момент закрытия торгов).

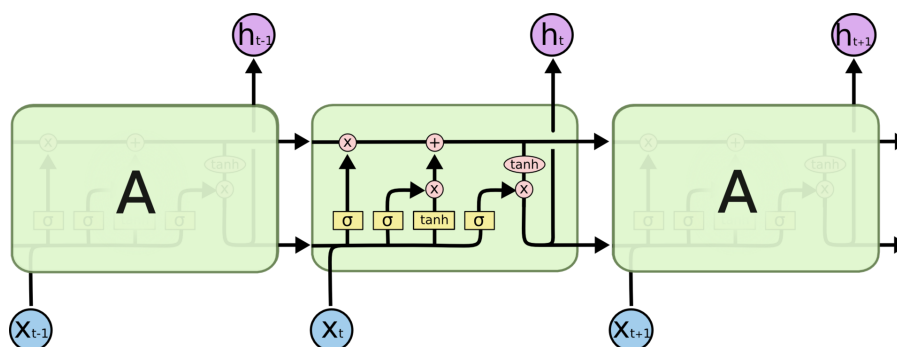


Рис. 1: Архитектура LSTM

II Планирование тестирования Разобьем данные о стоимости акций в отношении 80%/20% где 80 % - train 20% - test. На первом

наборе мы будем тренировать нашу модель, на тесте мы будем тестировать показатели. Оптимизировать будем следующие параметры: `lr` (скорость обучения), `weight_decay` (регуляризация), `look_back` (память), `hidden_dim` (количество узлов в скрытых слоях), `num_layers` (количество слоев). Финальная модель, которую мы будем тестировать [2](#)

```
LSTM(  
    (lstm1): LSTM(1, 64, batch_first=True)  
    (fc): Linear(in_features=64, out_features=1, bias=True)  
)
```

Рис. 2: Архитектура используемой нейросети

III Обучение модели.

- i. Тренировка модели под акции Apple. В процессе обучения выбрали следующие параметры исходя из тестирования модели, изначально коэффициент регуляризации был взят 10^{-5} как хороший бэкгранд, но лучшие показатели были при 10^{-6} . График падения `loss`-а изображен ниже [3](#) из него можно выделить то, что нейронная сеть обучается и градиент не "глохнет" в локальных минимумах.

<code>lr</code>	10^{-3}
<code>weight_decay</code>	10^{-6}
<code>look_back</code>	30
<code>hidden_dim</code>	64
<code>num_layers</code>	1
<code>epoch</code>	1000
<code>optimizer</code>	Adam

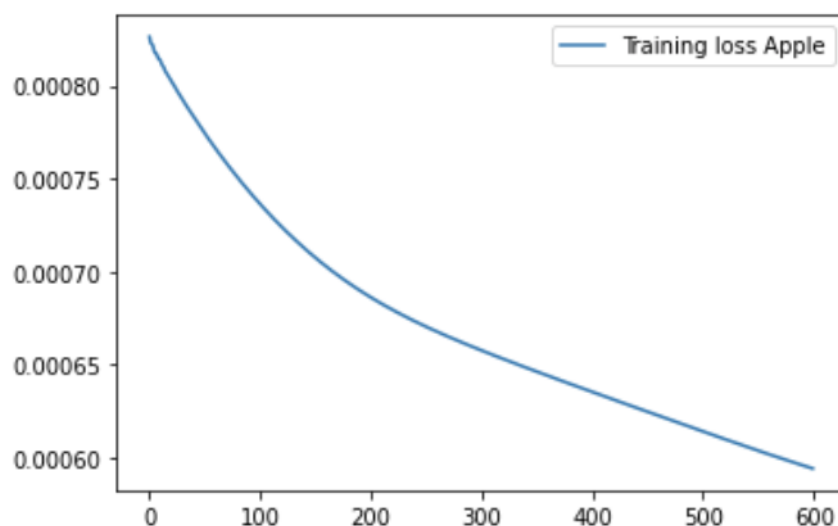


Рис. 3: Зависимость loss от эпохи

- ii. Тренировка модели под акции Nike. В процессе обучения выбрали следующие параметры.

lr	10^{-3}
weight_decay	10^{-5}
look_back	30
hidden_dim	64
num_layers	1
epoch	1000
optimizer	Adam

С ними изначально нейросеть показала хорошие результаты, поэтому мы их оставили без изменения. loss изображен ниже 3 Он тоже падает => нейросеть обучается и градиент сходится.

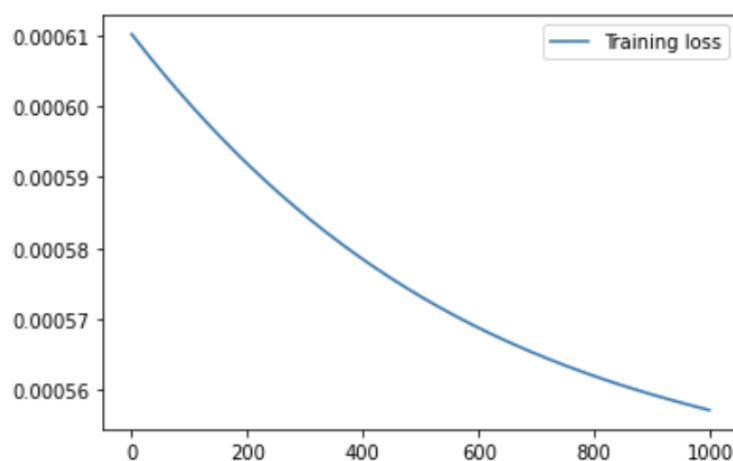


Рис. 4: Зависимость loss от эпохи для Nike

- iii. Тренировка модели под акции ExxonMobile. В процессе обучения выбрали следующие параметры.Здесь многое пришлось подкорректировать от изначальных. Было произведено динамическое обучение с разными показателями скорость (lr), также пришлось изменить look_back до 10. На графике ниже изображен loss на последних 200 эпохах. 5:

lr	$10^{-3} - 10^{-4}$
weight_decay	10^{-5}
look_back	10
hidden_dim	64
num_layers	1
epoch	1000
optimizer	Adam

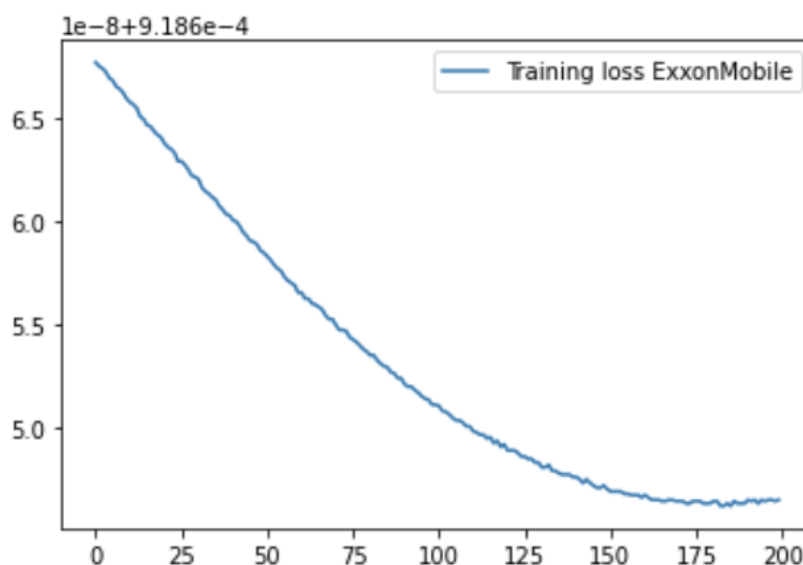


Рис. 5: Зависимость loss от эпохи для ExxonMobil

4. Оценка результатов.

I Метрики качества прогнозирования модели для акций Apple получились следующие:

Train RMSE	1.29
Train R^2	0.99
Test RMSE	1.91
Test R^2	0.87

Получилось добиться отличных показателей как на Train так и на Test выборке, это свидетельствует о том, что нейронная сеть обучилась на данных и с высокой точностью может предсказывать дальнейшую стоимость акций данной компании. С графиками предсказаний вы можете ознакомиться ниже.

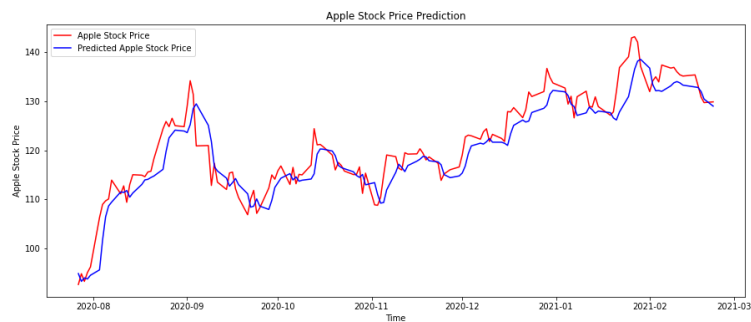


Рис. 6: Предсказания стоимости акций Apple на test

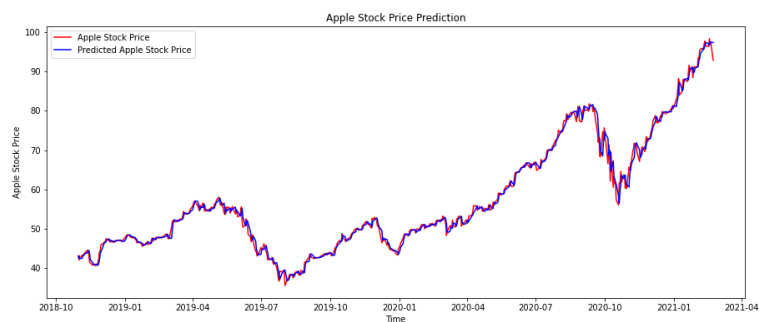


Рис. 7: Предсказания стоимости акций Apple на train

pass

II Метрики качества прогнозирования модели для акций Nike получились следующие.

Train RMSE	0.02
Train R^2	0.99
Test RMSE	0.33
Test R^2	0.94

Результаты получились потрясающие, среднеквадратичная ошибка (RMSE) на тесте составляет 0.33, хотя изначально цель была 0.8, $R^2 = 0.94$ получился потрясающий коэффициент детерминации, что свидетельствует о хорошей дисперсии предсказаний. Визуализацию предсказаний модели вы можете посмотреть ниже.

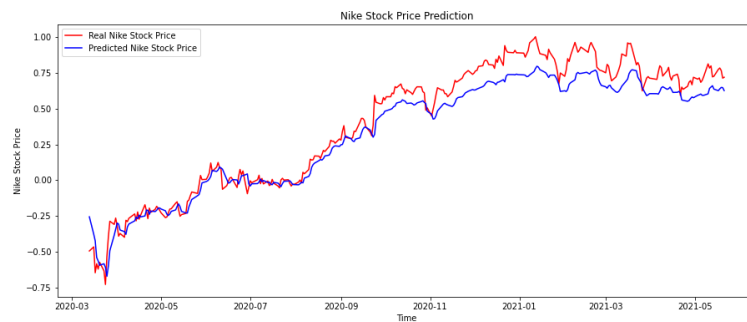


Рис. 8: Предсказания стоимости акций Nike на test

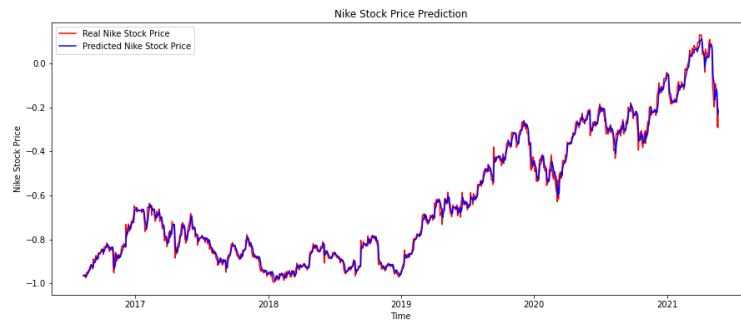


Рис. 9: Предсказания стоимости акций Nike на train

III Метрики качества прогнозирования модели для акций ExxonMobile получились следующие:

Train RMSE	0.96
Train R^2	0.98
Test RMSE	1.24
Test R^2	0.96

Точность предсказания на test замечательная, особенно хочу выделить коэффициент детерминации $R^2 = 0.96$, он намного выше изначальной цели, графики предсказаний ниже.

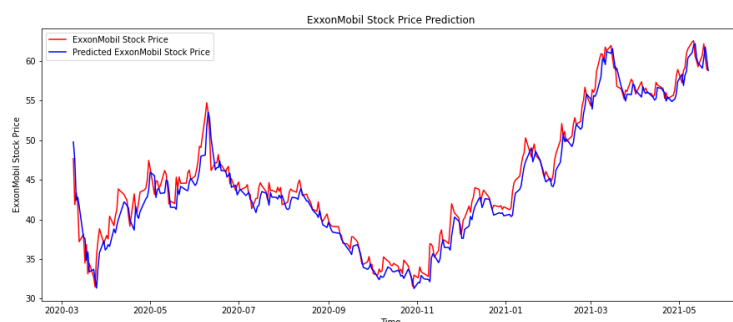


Рис. 10: Предсказания стоимости акций ExxonMobil на test

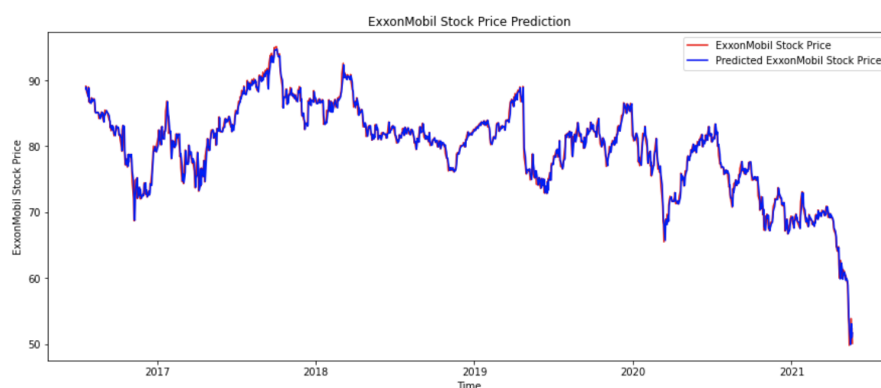


Рис. 11: Предсказания стоимости акций ExxonMobil на train

5. Вывод.

Таким образом RNN являются мощным инструментом для задач регрессии, мы получили очень хорошие стартовые показатели прогнозирования для нашей модели. Но в любом случае это только отправная точка для создания полной системы предсказания цен на акции, лучшую точность мы получили на акциях ExxonMobile, это говорит о том, что стоит подобрать те акции, для которых нейросеть предсказывает лучше, и делать систему именно для них. Также для улучшения точности можно использовать сглаживания рядов, и более сложные технологии. Но выбранная архитектура справляется с изначально поставленной задачей очень хорошо => добились чего хотели . Ссылка на [google colab](#)