

Прогнозирование курса котировок фонда S&P500 с помощью LSTM

Выполнил студент 206 группы
Физического факультета
Клиентов Григорий Алексеевич

Цель работы

- Исследовать поведение LSTM-модели на задаче прогнозирования курса фонда S&P500
- Проанализировать результаты в зависимости от внешних параметров модели LSTM

Описание реализации

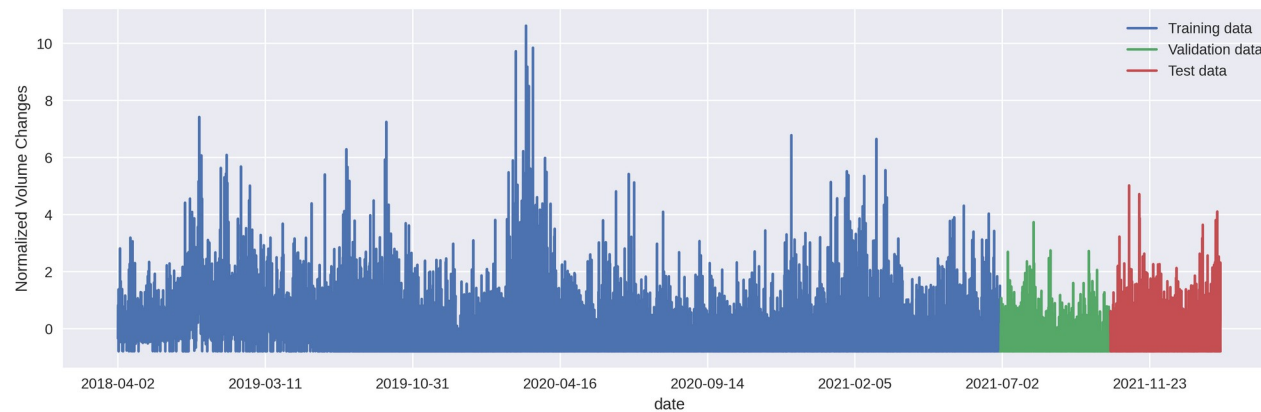
- Для имплементации модели была использована библиотека Tensorflow
- Датасет взят из Tinkoff Invest Api за период 2018-04 по 2021-04 с шагом свечи в 1 час
- На вход модели подаётся вся информация свечи, со специфическим учетом даты
- Требуем от модели прогноз курса закрытия свечек
- Обучение проводим на 50-ти эпохах

Идея исследования

- Обученную модель заставим предсказывать предстоящий курс за 200 часов
- В дате свечи учитывается день недели и время закрытия, обёртывая в синус
- Полученное предсказание будем оценивать так:
 1. Разделим истинный и предсказанный курсы на участки по 7 часов и отметим те, на которых колебания курса больше 0.5%
 2. Посмотрим на совпадение и несовпадение таких отметок на истинной и предсказанном курсах

Исходный датасет

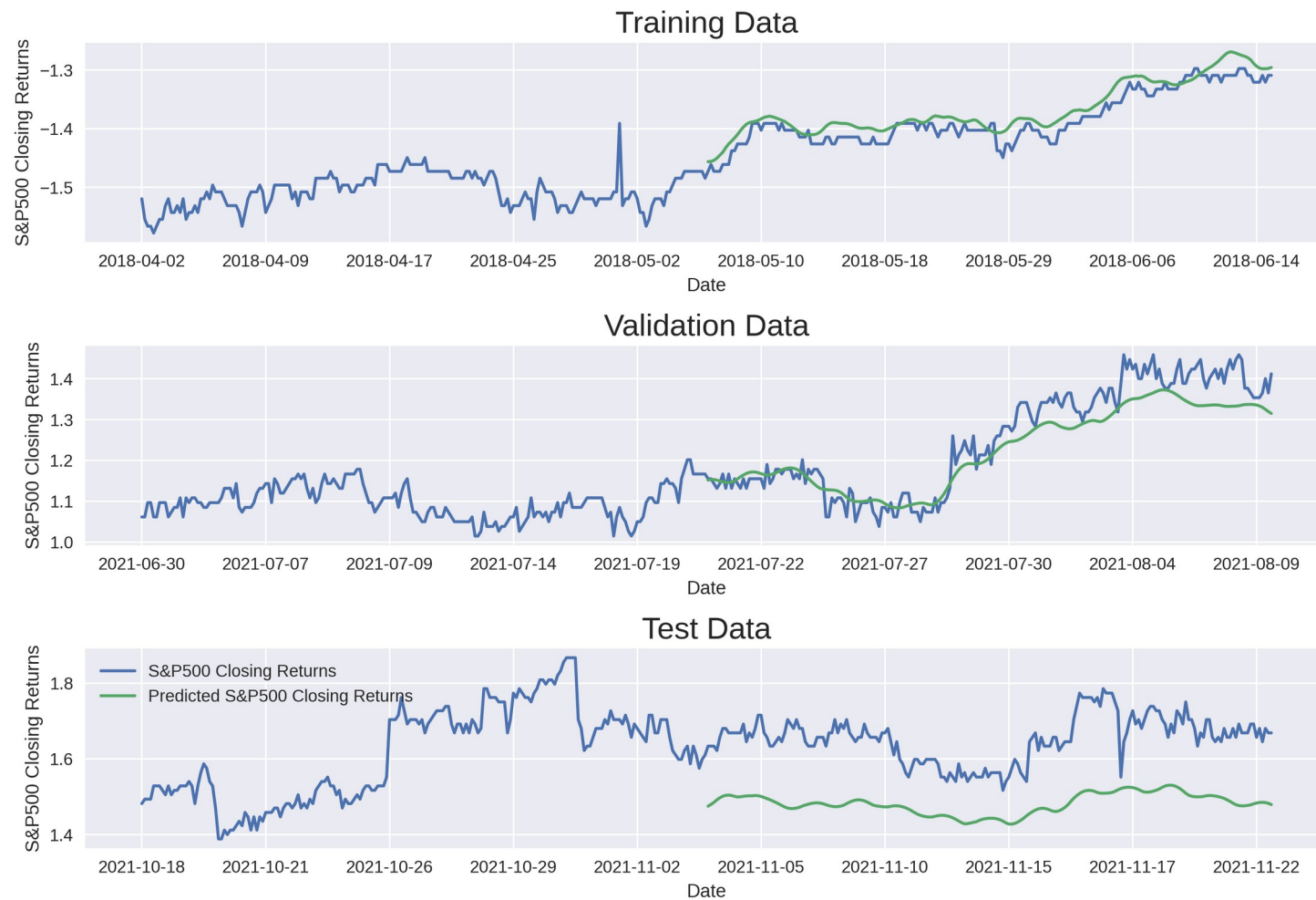
Data Separation



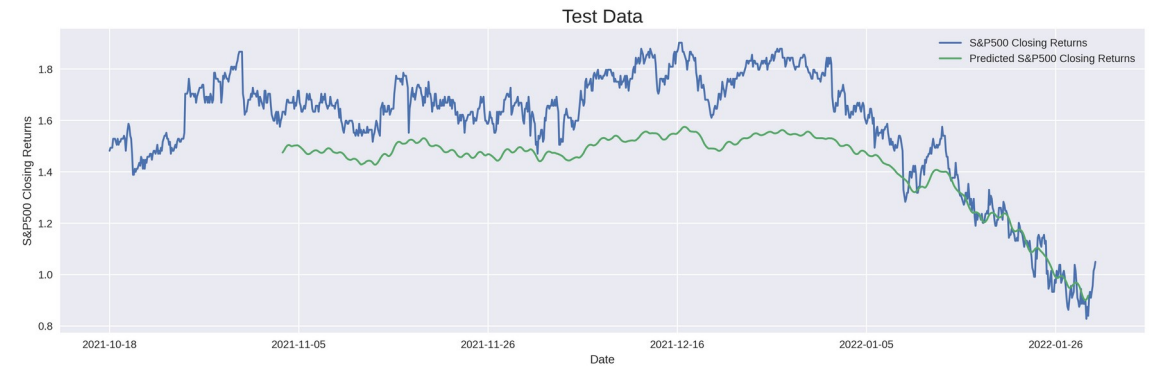
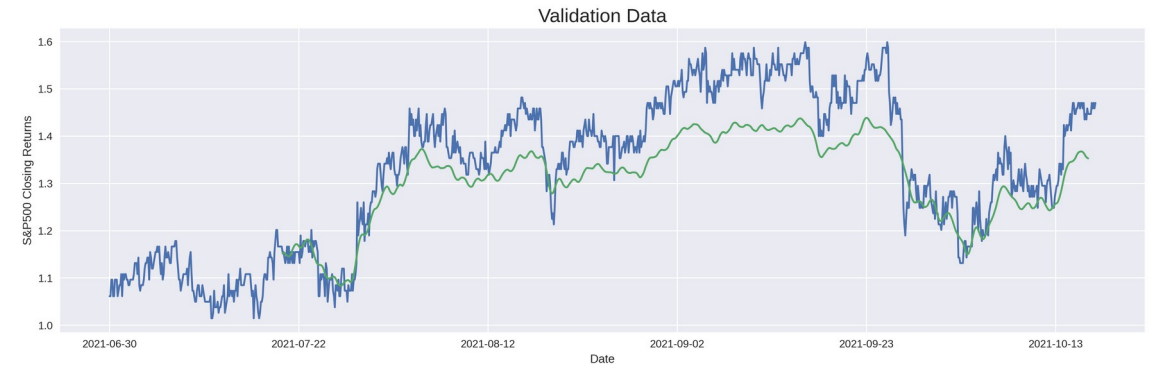
Результаты

Проведем исследование предсказаний при варьировании параметров длины обучающей последовательности, а также от наличия учета времени в наборе данных.

$SEQ_LEN = 200$
Дата учитывается



Предсказание на всём промежутке

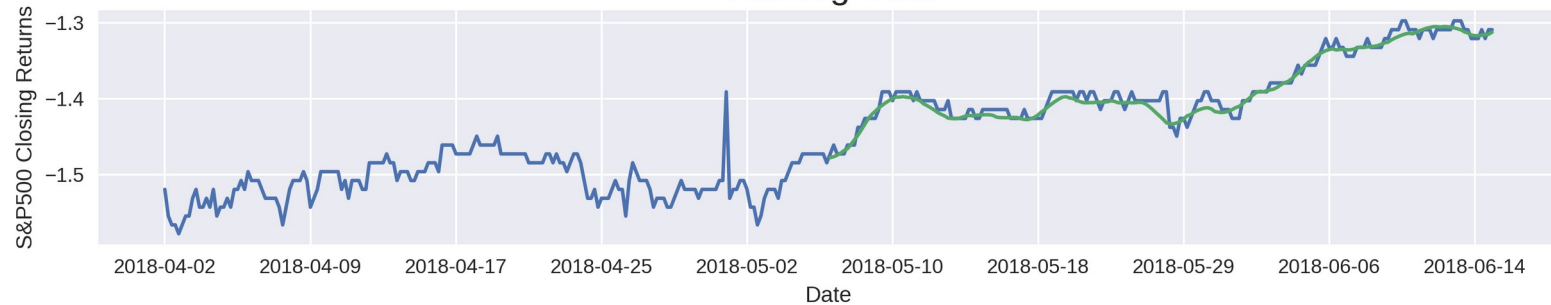


$SEQ_LEN = 200$

Дата **НЕ**

учитывается

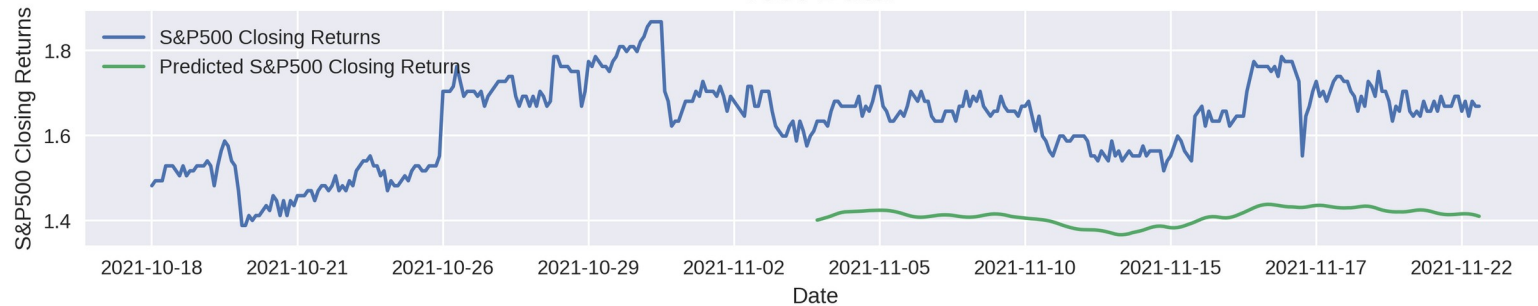
Training Data



Validation Data



Test Data



Предсказание на всём промежутке



Также было оценено количество дней, где колебания курса составляли $> 0.5\%$

	True days	Predict days	Conv days	Unconv days	Date consider
SEQ_LEN					
100	142	110	110	32	True
100	142	83	83	59	False
200	128	111	111	17	True
200	128	69	69	59	False
300	113	106	106	7	True
300	113	75	75	38	False

Основные итоги работы

- Было исследован результат модели для различных SEQ_LEN — визуально предсказания не отличались друг от друга
- Были исследован результат в зависимости от учета даты — с учетом даты предсказания получались более точные
- Было проанализировано качество предсказания на методе поиска дней с колебанием курса бумаги более заданного значения (0.5%).

Из таблицы видно, что наилучшие результаты показывали модели, учитывающие дату