

# Прогнозирование курса котировок фонда S&P500 с помощью LSTM

Выполнил студент 206 группы  
Физического факультета  
Клиентов Григорий Алексеевич

# Цель работы

- Предсказать курс фонда S&P500 с помощью LSTM-модели
- Проанализировать результаты в зависимости от длины входящей последовательности и от учета даты

# Описание реализации

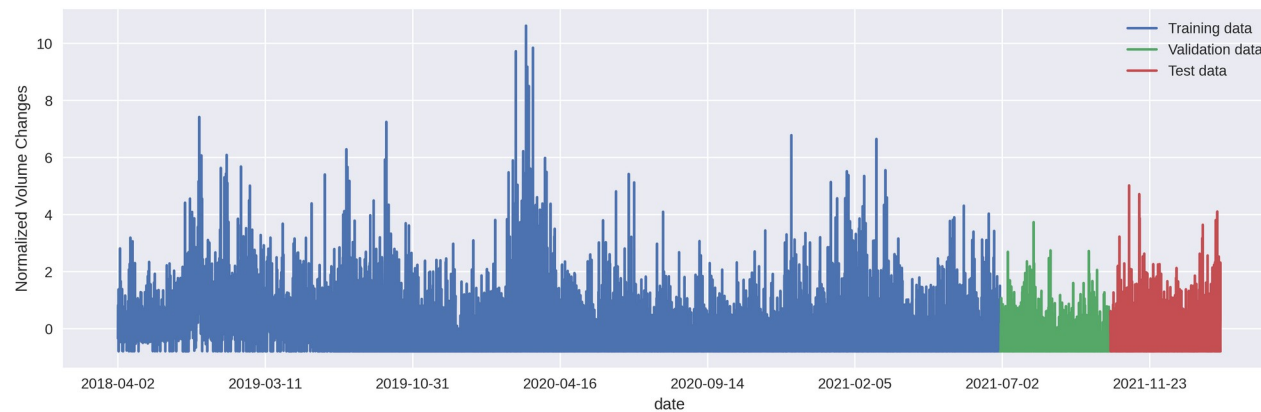
- Для имплементации модели была использована библиотека Tensorflow
- Датасет взят из Tinkoff Invest Api за период 2018-04 по 2021-04 с шагом свечки в 1 час
- На вход модели подаётся вся информация свечки, со специфическим учетом даты
- Обучение проводим на 50-ти эпохах
- Модель содержит 4 слоя:
  - I. *LSTM — 64 выходная размерность,*
  - II. *LSTM — 64 выходная размерность*
  - III. *LSTM — 32 выходная размерность*
  - IV. *Dense — без функций активации*

# 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

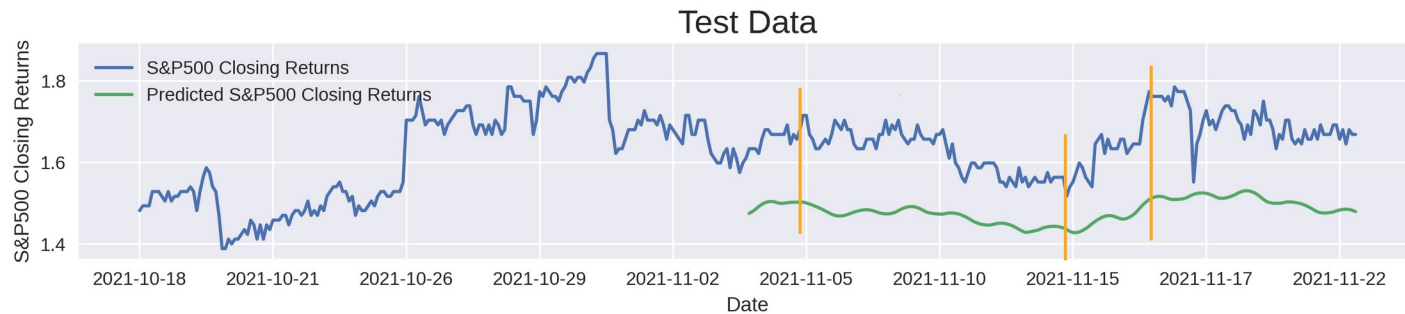
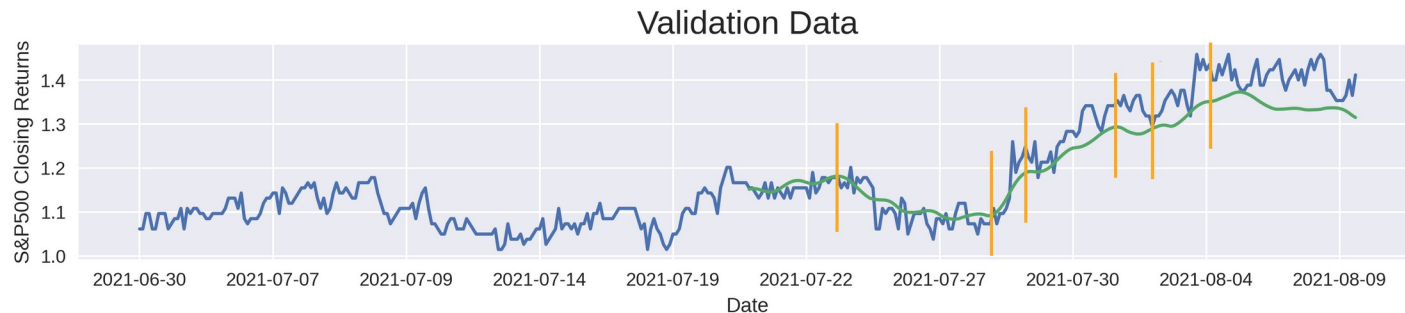
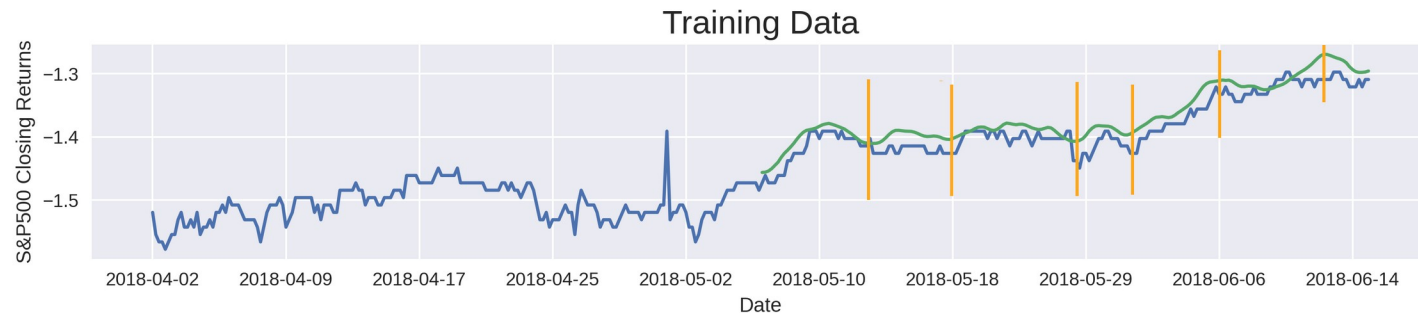
Data Separation



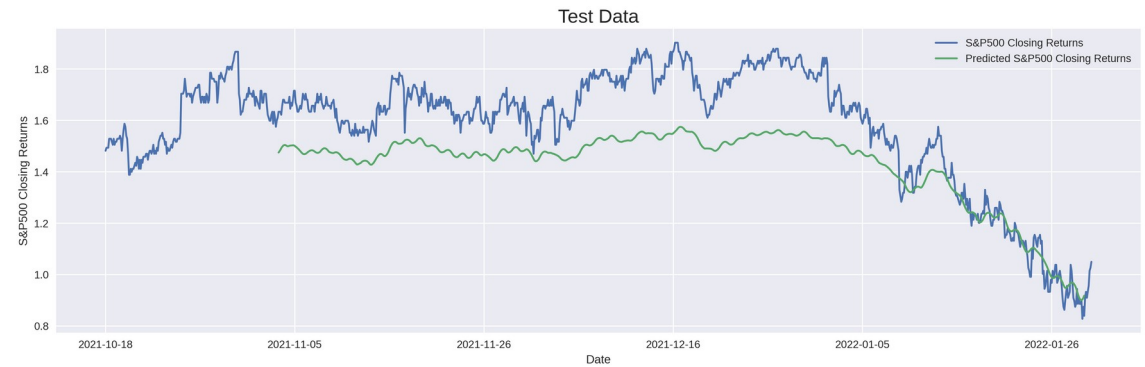
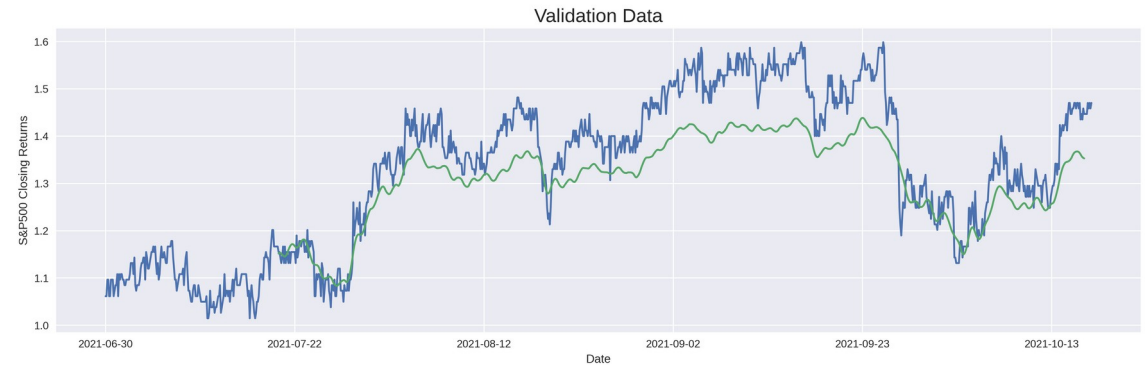
# Результаты

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

- $SEQ\_LEN = 200$
- Дата учитывается

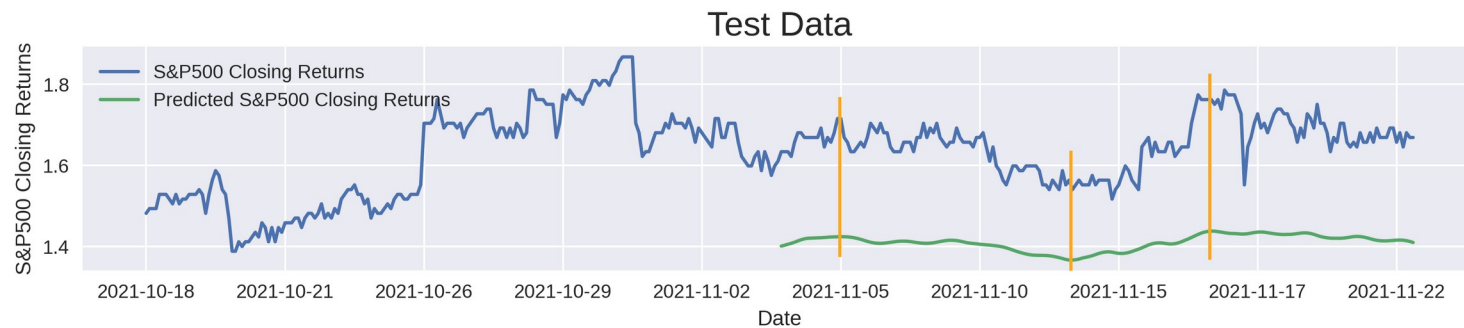
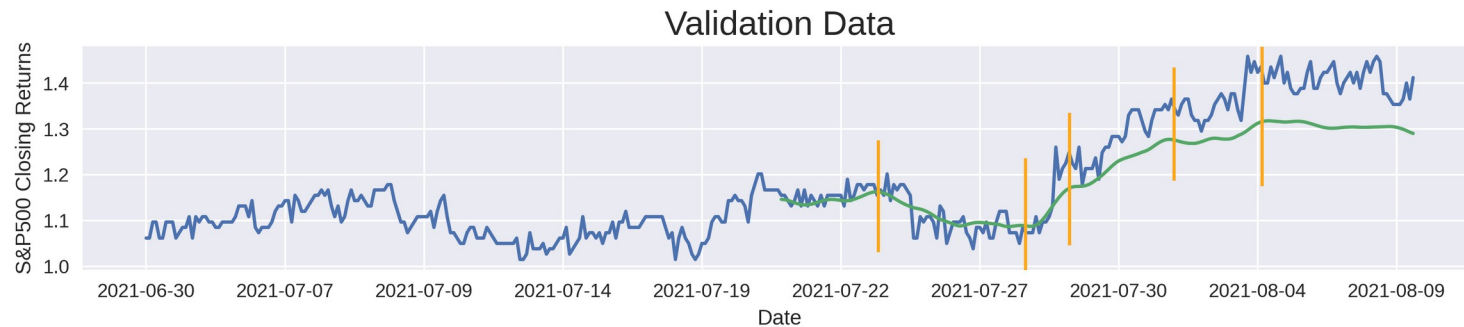
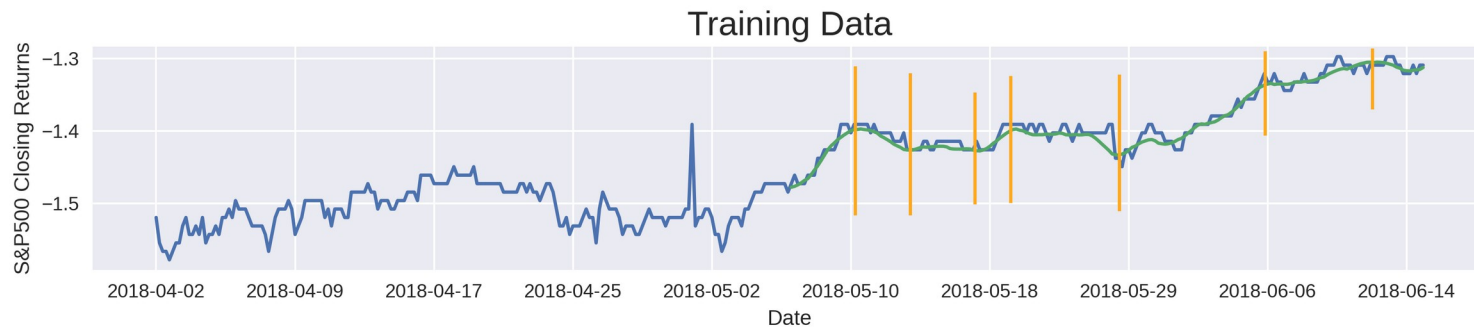


# Предсказание на всём промежутке





- $SEQ\_LEN = 200$
- Дата НЕ
- учитывается



# 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

Видно, что при прочих равных модель, которая учитывает дату, выдаёт лучшие результаты.

# Основные итоги работы

- Была реализована модель LSTM для прогнозирования курса фонда S\$P500
- Был исследован результат в зависимости от учета даты во входных данных. Качество предсказания повышалось при учете даты при прочих равных условиях.
- Была опробована новая метрика качества предсказания — совпадения дней с колебанием курса более заданного значения ( $>0.5\%$ ). Модели, учитывающие даты, показали лучший результат по сравнению с неучитывающими.
- Также, было проведено исследование качества предсказания моделей от длины входной последовательности.