

Исследование эффективности нейронных сетей для прогнозирования финансовых временных рядов

Хорунженко Аркадий Сергеевич

Группа 22712

Научный руководитель: **к.э.н. Макушев Василий Леонидович**

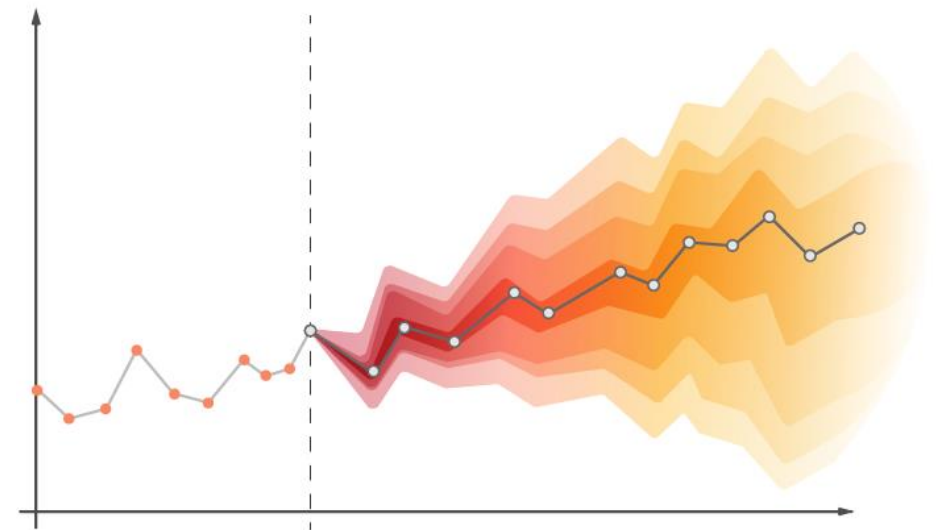
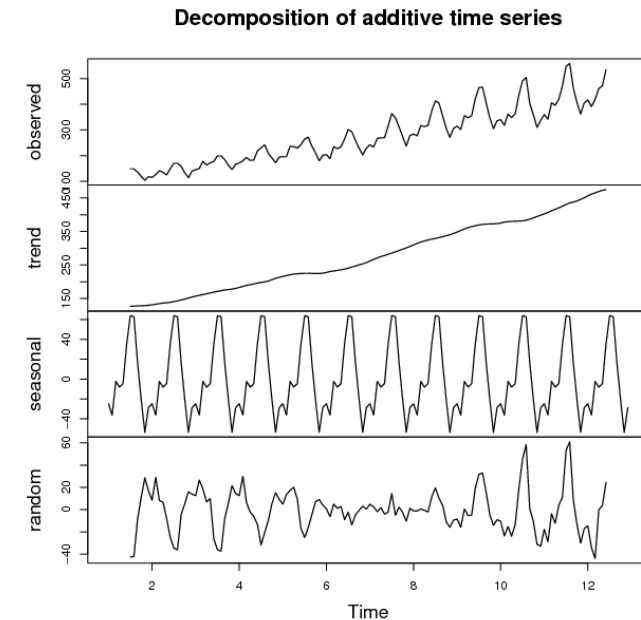
Рецензент: на данный момент отсутствует

Новосибирский Государственный университет

- Экономика и финансы: принятие решений в области инвестиций
- Медицина: прогнозирование заболеваемости
- Энергетика: планирование и управление энергосистемами
- Логистика: управление трафиком, прогнозирование пробок



- Прогнозирование позволяет автоматизировать и оптимизировать процессы принятия решений на основе анализа изменения показателей во времени.
- Нейронные сети, как один из методов искусственного интеллекта, позволяют строить сложные модели, учитывающие множество факторов, что улучшает точность прогнозов и позволяет решать более сложные задачи.



- Цель - исследовать эффективность нейронных сетей для прогнозирования временных рядов с учётом различной разряженности экономических данных, а также в сравнительном анализе с традиционными методами прогнозирования.

Цели и задачи

- Задачи:

1. Провести обзор литературы и анализ существующих методов прогнозирования временных рядов.
2. Изучить информацию о статистических методах и методах машинного обучения.
3. Подготовить данные для прогнозирования
4. Подобрать оптимальные параметры для различных архитектур нейронных сетей и методов обучения на основе анализа результатов экспериментов с использованием различных наборов данных.
5. Разработать и апробировать модель для прогнозирования различных временных рядов и оценить ее эффективность по сравнению с традиционными методами, а также сравнить эффективность разработанных моделей на основе различных критериев качества прогнозирования.
6. Проанализировать применимость различных архитектур и методов обучения нейронных сетей для прогнозирования различных видов временных рядов, и оценить их преимущества и недостатки.
7. Сделать выводы о эффективности нейронных сетей для прогнозирования временных рядов и их эффективности, а также о возможных направлениях дальнейших исследований в области прогнозирования временных рядов.

1. Проанализирована литература по теме исследования

2.1. Изучены различные статистические методы прогнозирования временных рядов

$$X_t = c + \sum_{i=1}^p a_i X_{t-i} + \varepsilon_t$$

Авторегрессионная модель

$$X_t = c + \sum_{i=1}^p a_i X_{t-i} + \sum_{i=1}^q b_i \varepsilon_{t-i} + \varepsilon_t$$

ARMA

$$\Delta^d X_t = c + \sum_{i=1}^p a_i X_{t-i} + \sum_{i=1}^q b_i \varepsilon_{t-i} + \varepsilon_t$$

ARIMA

$$x_t = \sigma_t e_t$$
$$\sigma_t^2 = a_0 + a_1 x_{t-1}^2 + \dots + a_p x_{t-p}^2$$

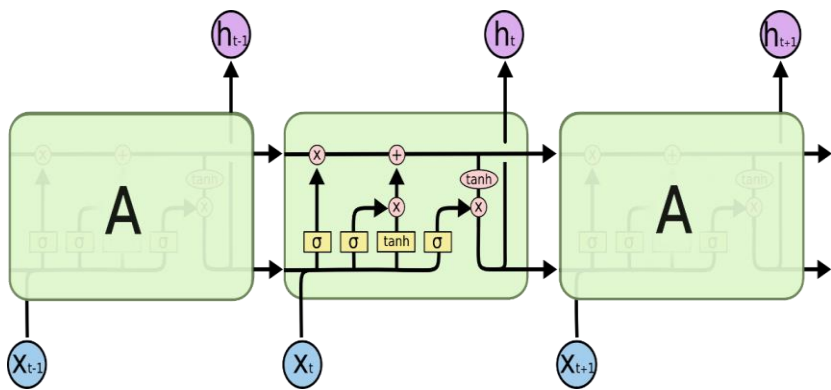
ARCH

$$\sigma_t^2 = \omega + \sum_{i=1}^q a_i u_{t-i}^2 + \sum_{j=1}^p b_j \sigma_{t-j}^2$$

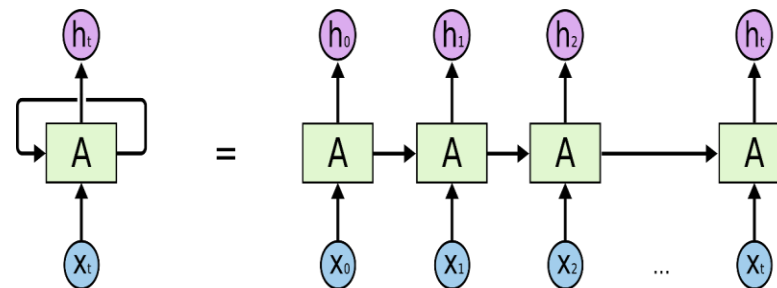
$$\sum_{i=1}^q a_i + \sum_{j=1}^p b_j < 1$$

GARCH

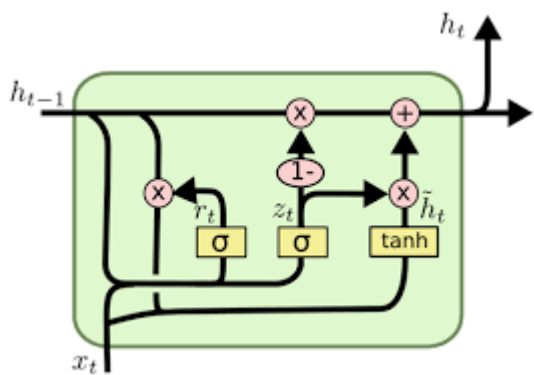
2.2. Изучены различные архитектуры нейронных сетей



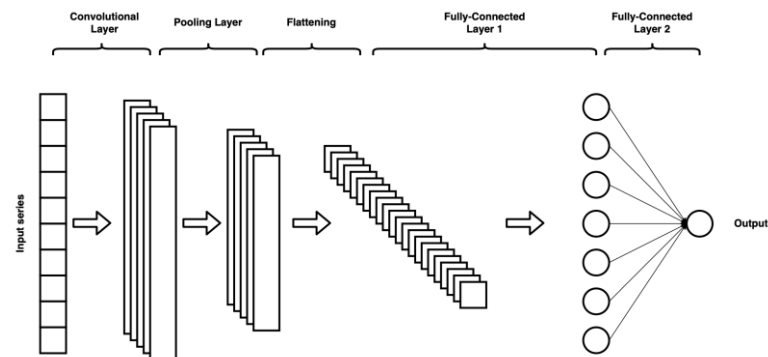
LSTM



RNN



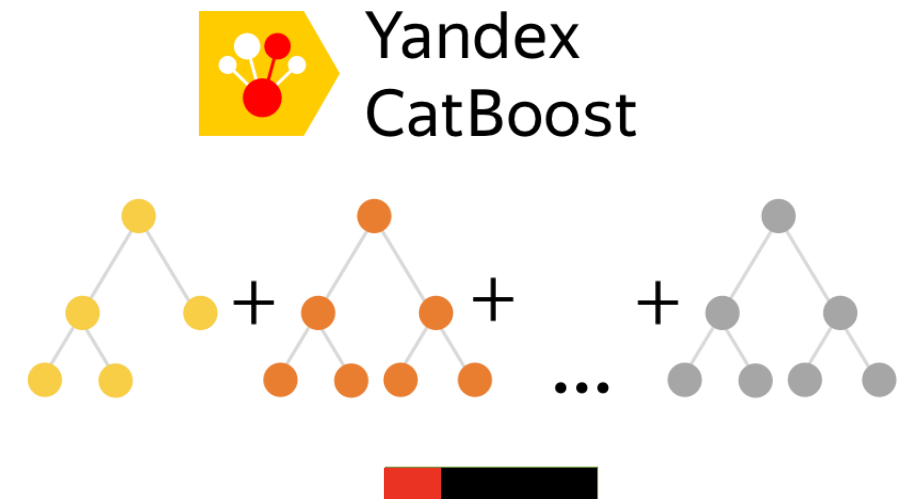
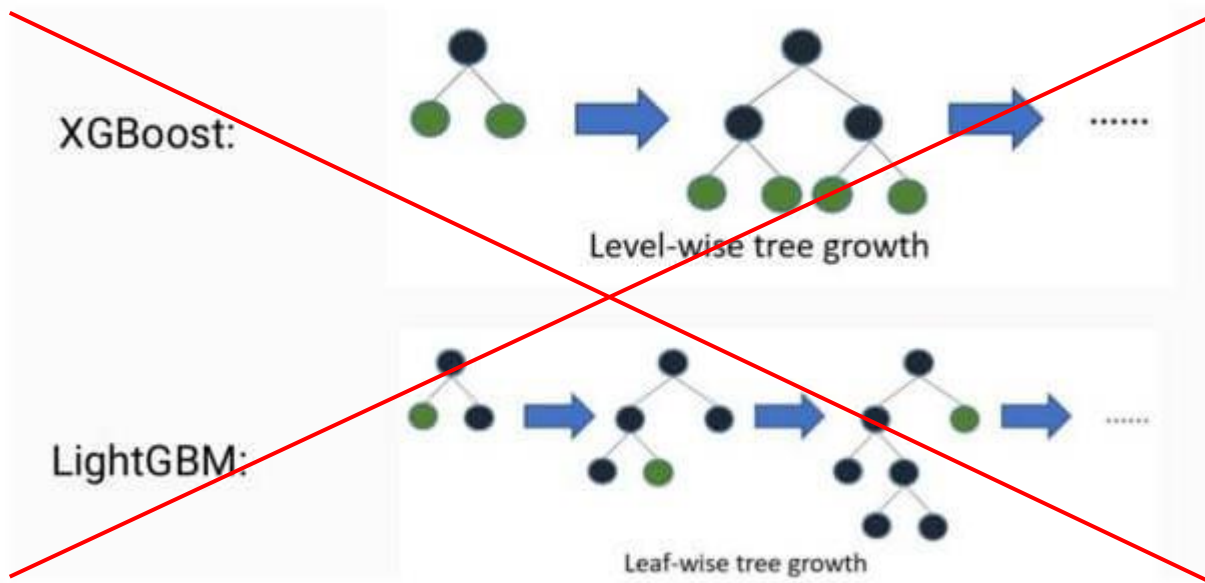
GRU



CNN

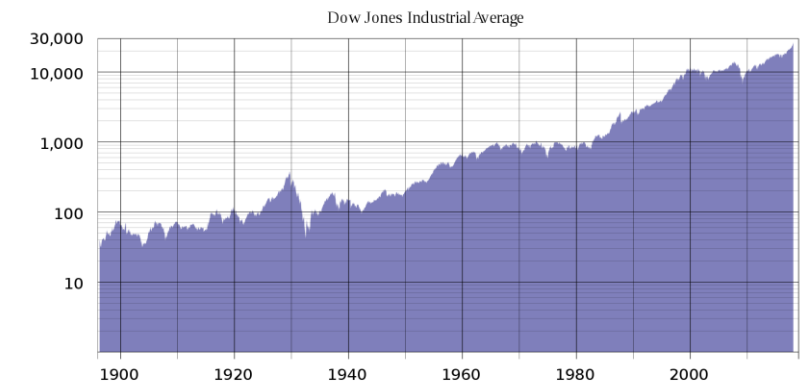
2.3. Изучены методы машинного обучения использующие градиентный бустинг

По результатам исследования было решено использовать только градиентный бустинг catboost

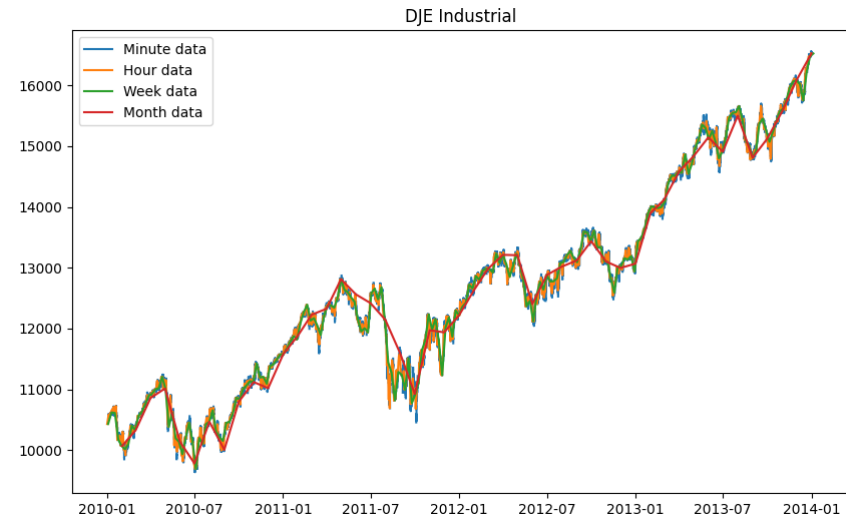
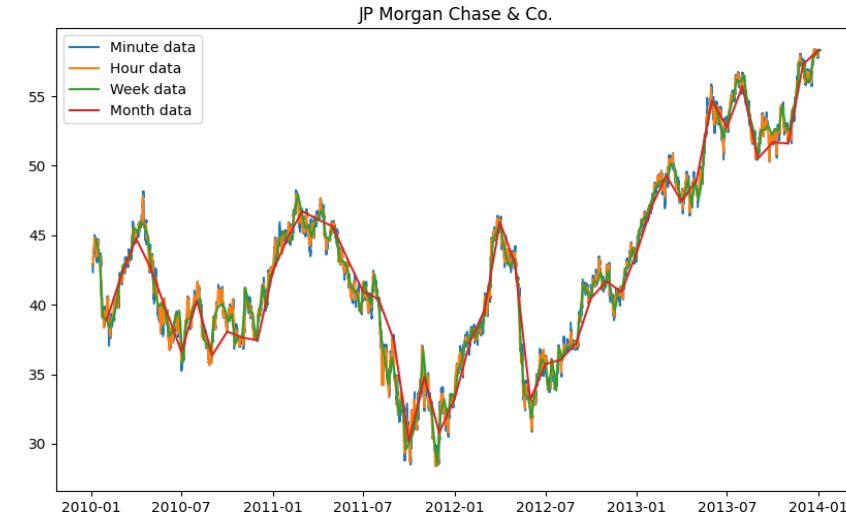
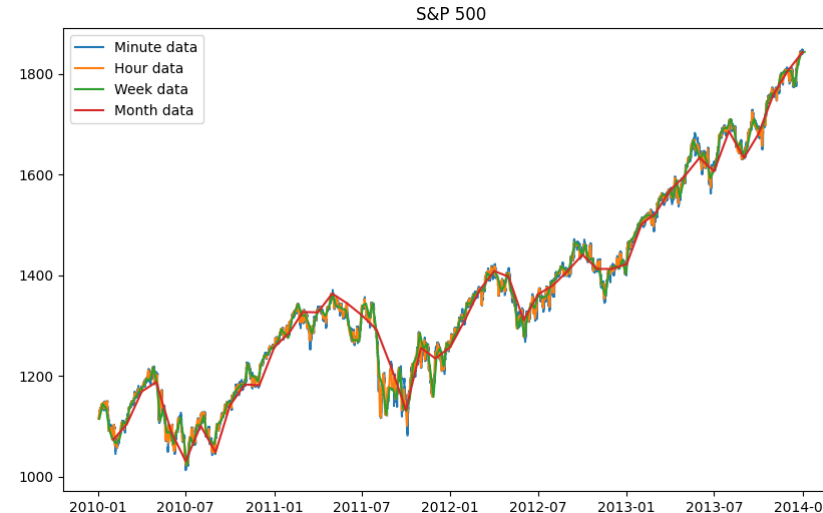


3. На данном этапе исследования взяты следующие наборы данных за период 2010 - 2015 г.г.:

- Фондовый индекс S&P500
- Промышленный индекс Dow Jones
- Акции JPMorgan Chase & Co
- Акции NASDAQ



3.1. Наборы были получены с минутным шагом и были преобразованы данные с разной частотой: минутные, часовые, дневные, недельные и месячные.



3.2. Для каждого набора данных был проведён статистический анализ. Пример для S&P500 часовых данных:

Сравнение показателей среднего арифметического (mean) и медианы (median) свидетельствует о правосторонней асимметрии (т.к. $\text{mean} > \text{median}$);

Значение коэффициента вариации свидетельствует об однородности исходных данных ($CV = 0.1462 < 0.33$);

Значение показателя асимметрии skew (As) свидетельствует об умеренной правосторонней асимметрии ($As = 0.4497$, $|As| < 0.5$, $As > 0$);

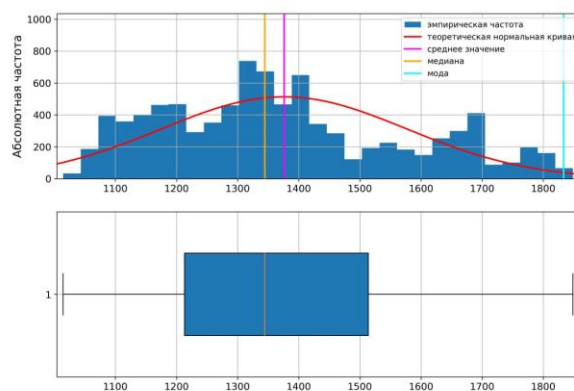
Значение показателя асимметрии kurtosis (Es) свидетельствует о плосковершинном распределении ($Es = -0.694$);

Коробчатая диаграмма показывает отсутствие аномальных значений (выбросов) для всей совокупности.

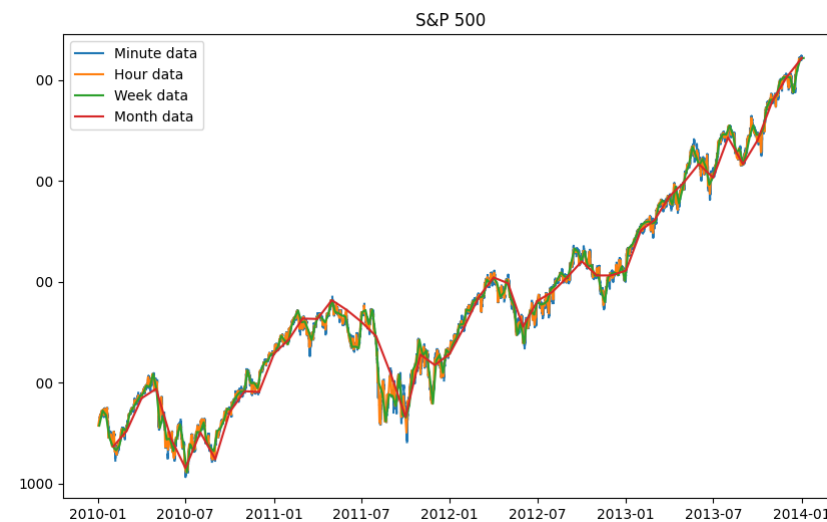
Вероятностные графики свидетельствует о том, что скорее всего закон распределения отличается от нормального.

```
descriptive_characteristics(np.array(data['Close']), p_level=0.95)
```

characteristic	evaluation	conf.int.low	conf.int.high	abs.err.	rel.err.(%)	note
0	count	9010.000000				
1	mean	1376.018632	1371.862859 1380.174404	2.120046	0.154071	
2	median	1344.305000	1341.47 1348.54	2.657084	0.197655	distribution is positive skewed (правостороння...
3	mode	1833.320000				
4	variance	40496.312624	39339.246399 41705.436071	301.674143	0.744942	
5	standard deviation	201.236956	198.341237 204.219088	1.499099	0.744942	
6	mean absolute deviation	163.938950				
7	min	1014.680000				
8	5%	1090.404000				
9	25% (Q1)	1212.947500				
10	50% (median)	1344.305000				
11	75% (Q3)	1513.647500				
12	95%	1756.584000				
13	max	1848.360000				
14	range = max - min	833.680000				
15	IQR = Q3 - Q1	300.700000				
16	CV = mean/std	0.146238	0.14443 0.148091	0.001113	0.760749	CV <= 0.33 (homogeneous population)
17	QCD = (Q3-Q1)/(Q3+Q1)	0.110284				
18	skew (As)	0.449726		0.025801	5.737113	distribution is moderately positive skewed (ум...
19	kurtosis (Es)	-0.694200		0.051597	-7.432552	platykurtic distribution (плосковершинное расп...



коробчатая диаграмма и гистограмма
распределения



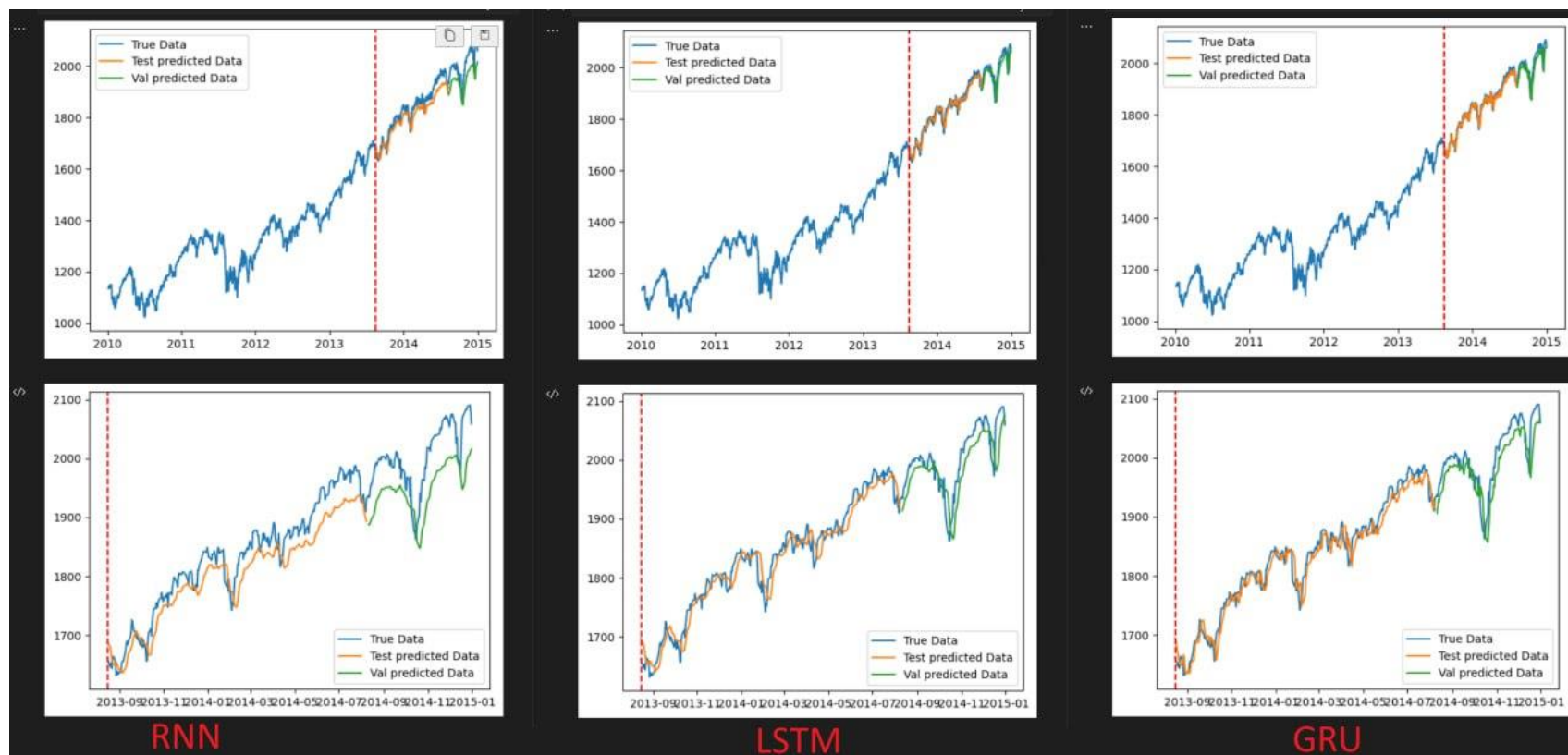
3.2. Наборы данных имеют пропуски в выходные и праздничные дни. В зависимости от варианта заполнения пропусков получаем ещё несколько наборов данных:

1. Сдвиг даты
2. Заполнение последним рабочим днём
3. Линейное заполнение между двумя рабочими днями
4. Использование категориальных признаков (Для подходов DL и ML)

4. Получены результаты для датасета S&P500(2010-2015) для моделей нейронных сетей

Весь датасет с
участком
предсказания и
теста

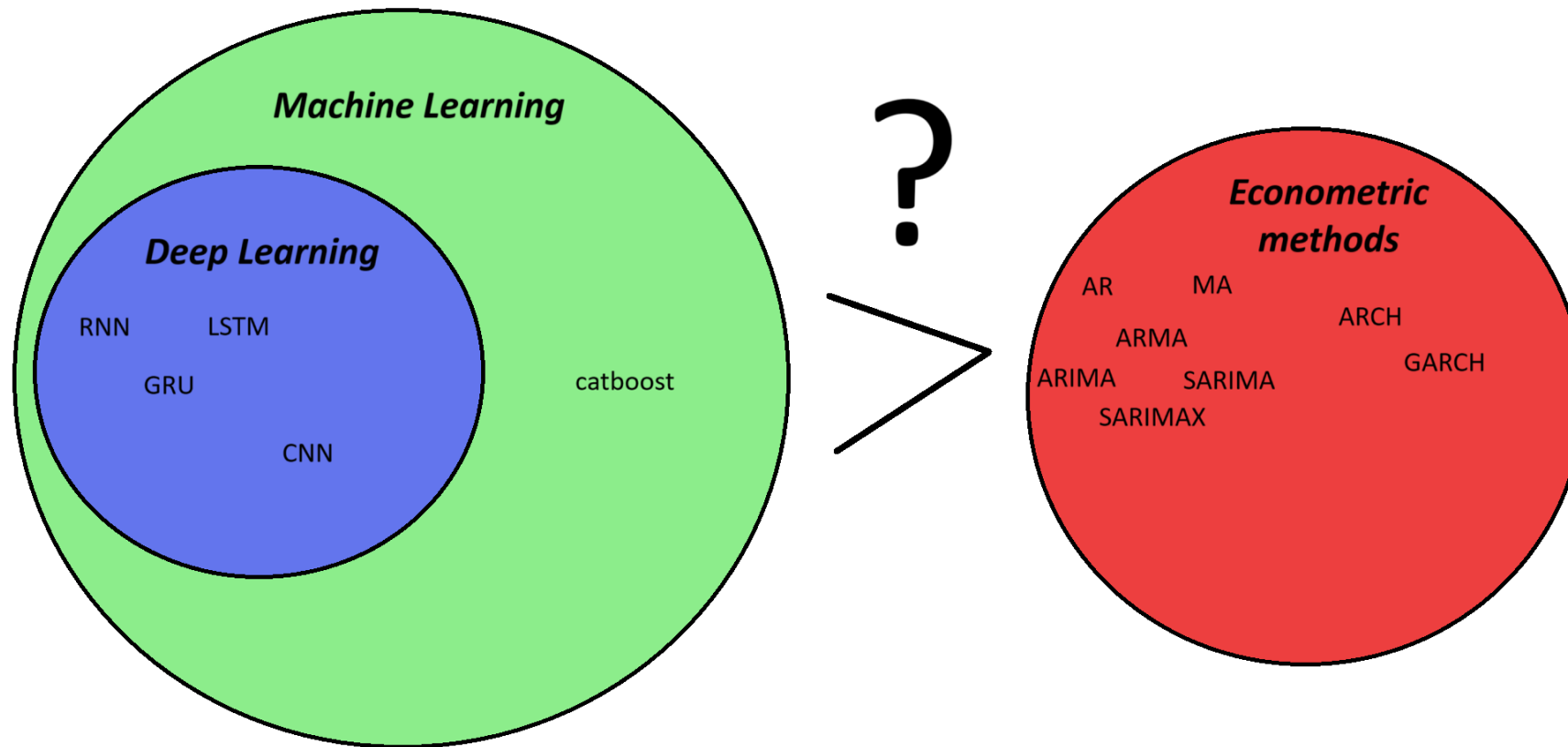
Предсказание:
Тест и валидация



1. Реализованы и обучены модели нейронных сетей с архитектурами RNN, LSTM, GRU
2. Реализована модель нейронных сетей CNN, но пока с данной моделью получены не достаточно удовлетворительные результаты
3. Реализованы традиционные методы прогнозирования такие как AR, MA, ARMA, ARIMA, SARIMA, SARIMAX, ARCH и GARCH
4. Увеличено количество наборов данных
5. Реализован код для работы с моделью градиентного бустинга catboost
6. Начато написание текста ВКР
7. Реализованный код оформлен в виде jupyter notebook'ов и в виде библиотеки

1. Сравнить эффективность разработанных моделей и традиционных методов прогнозирования на основе различных критериев качества прогнозирования
2. Проанализировать применимость различных архитектур и методов обучения нейронных сетей для прогнозирования различных видов временных рядов и оценить их преимущества и недостатки

3. Сделать выводы о применимости нейронных сетей для прогнозирования временных рядов и их эффективности по сравнению с традиционными методами, а также о возможных направлениях дальнейших исследований в области прогнозирования временных рядов с использованием нейронных сетей.



Отсутствуют на данный момент

Апробация:

- Планируется участие в конференции.

Список литературы

1. https://www.finam.ru/profile/fyuchersy-usa/nq-100-fut/export/?market=7&em=21719&token=&code=NDX&apply=0&df=1&mf=0&yf=2010&from=01.01.2010&dt=1&mt=0&yt=2014&to=01.01.2014&p=2&f=NDX_100101_140101&e=.txt&cn=NDX&dtf=1&tmf=1&MSOR=1&mstime=on&mstimever=1&sep=1&sep2=1&datf=1&at=1
2. https://www.finam.ru/profile/fyuchersy-usa/sandp-fut/export/?market=7&em=108&token=&code=SP&apply=0&df=1&mf=0&yf=2010&from=01.01.2010&dt=1&mt=0&yt=2014&to=01.01.2014&p=2&f=SP_100101_140101&e=.txt&cn=SP&dtf=1&tmf=1&MSOR=1&mstime=on&mstimever=1&sep=1&sep2=1&datf=1&at=1
3. Engle, Robert F, Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. 1982, p. 987–1007, Econometrica.
4. Bollerslev, Tim, Generalized Autoregressive Conditional Heteroskedasticity, 1986, p. 307–327, Journal of Econometrics.
5. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
6. https://neerc.ifmo.ru/wiki/index.php?title=%D0%A0%D0%B5%D0%BA%D1%83%D1%80%D1%80%D0%B5%D0%BD%D1%82%D0%BD%D1%8B%D0%B5_%D0%BD%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D1%8B%D0%B5_%D1%81%D0%B5%D1%82%D0%B8
7. <https://github.com/SkivHisink/MasterWork>
8. <https://web.archive.org/web/20211110112626/http://www.wildml.com/2015/10/recurrent-neural-network-tutorial-part-4-implementing-a-grulstm-rnn-with-python-and-theano/>
9. <https://arxiv.org/abs/1412.3555>
10. https://www.sciencedirect.com/science/article/pii/S2212827121003796?ref=pdf_download&fr=RR-2&rr=79fdf5739bb63aad
11. “Лекционные и методические материалы”, Канторович Г.Г.,
12. Time Series Forecasting in Python, Marco Peixeiro, ISBN: 97816172998896 2022г.
13. Practical Time Series Analysis, Aileen Nielsen, ISBN: 9781492041658
14. Introduction to Time Series and Forecasting Second Edition, Peter J. Brockwell Richard A. Davis, ISBN 0387953515, 2002 г.
15. Machine Learning for Time Series Forecasting with Python, Francesca Lazeri, ISBN: 9781119682363, 2021г.
16. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. - М.: ФИЗМАТЛИТ, 2006. - 816 с.
17. Львовский Е.Н. Статистические методы построения эмпирических формул. - М.: Высшая школа, 1988. - 239 с.
18. Фёрстер Э., Рёнд Б. Методы корреляционного и регрессионного анализа / пер с нем. - М.: Финансы и статистика, 1983. - 302 с.
19. Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ / пер с англ. - М.: Мир, 1982. - 488 с.
20. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Книга 1 / пер.с англ. - М.: Финансы и статистика, 1986. - 366 с.
21. Айвазян С.А. и др. Прикладная статистика: Исследование зависимостей. - М.: Финансы и статистика, 1985. - 487 с.
22. Прикладная статистика. Основы эконометрики: В 2 т. 2-е изд., испр. - Т.2: Айвазян С.А. Основы эконометрики. - М.: ЮНИТИ-ДАНА, 2001. - 432 с.
23. Магнус Я.Р. и др. Эконометрика. Начальный курс - М.: Дело, 2004. - 576 с.
24. Носко В.П. Эконометрика. Книга 1. - М.: Издательский дом "Дело" РАНХиГС, 2011. - 672 с.
25. Брюс П. Практическая статистика для специалистов Data Science / пер. с англ. - СПб.: БХВ-Петербург, 2018. - 304 с.
26. Уатт Дж. и др. Машинное обучение: основы, алгоритмы и практика применения / пер. с англ. - СПб.: БХВ-Петербург, 2022. - 640 с.