

УДК 330.46

***ПРИМЕНЕНИЕ МЕТОДОВ АНАЛИЗА BIG DATA ДЛЯ
ПРОГНОЗИРОВАНИЯ СТОИМОСТИ ФИНАНСОВЫХ ИНСТРУМЕНТОВ***

Кучко А.Ю.,

студентка 2 курса бакалавриата

Самарский Государственный Экономический Университет

Самара, РФ

Кораблев А.В.,

к.э.н., доцент кафедры цифровых технологий и решений

Самарский Государственный Экономический Университет

Самара, РФ

Аннотация: В статье обосновываются преимущества использования методов обработки больших данных для прогнозирования тренда финансовых инструментов по сравнению с малыми выборками статистических данных. В ходе исследования проведен анализ котировок акций Caterpillar за 32 года с помощью библиотеки машинного обучения Python Stocker, внесены модификации в исходный код, сделаны выводы относительно эффективности использования подобных инструментов для принятия стратегических решений типа buy-hold.

Ключевые слова: большие данные, big data, python, stocker, прогнозирование, финансы.

***APPLICATION OF THE BIG DATA ANALYSIS METHODS FOR
PREDICTING THE COST OF FINANCIAL INSTRUMENTS***

Kuchko A.Yu.,

2nd year bachelor's student

Samara State University of Economics

Samara, Russia

Korablev A.V.,

Candidate of Economics, Associate Professor of Department of Digital Technologies and Solutions

Samara State University of Economics

Samara, Russia

Abstract: The article substantiates the advantages of using big data processing methods for predicting the trend of financial instruments compared to using small samples of statistical data. In the course of the study, Caterpillar stock quotes for 32 years were analyzed using the Python Stocker machine learning library. Modifications were made to the source code. Conclusions are based on the effectiveness of using such tools for making strategic decisions of the “buy-hold” type.

Key words: big data, python, stocker, prediction, finance.

Введение. С ростом возможностей по сбору, хранению и анализу больших объемов информации возникает объективная необходимость по разработке и внедрению наиболее эффективных инструментов обработки имеющихся данных. Дискуссионным остается вопрос о применении понятия big data («большие данные»). Расхождения в дефиниции возникают при определении конкретного объема данных, являющегося маркером для обозначения их «большими». Для целей данного исследования под big data будем понимать статистические данные за период, превышающий один год. Подобный подход обоснован тем, что в колебаниях котировок акций за день, неделю или месяц высока доля неслучайных экзогенных факторов. К ним можно отнести: время открытия и закрытия торгов, продолжительность рабочей недели на бирже, количество выходных и праздничных дней в каждый конкретный месяц и т.п. В то же время на годовом горизонте есть возможность наблюдать сезонность

временных рядов (порой ярко выраженную), кроме того, более отчетливо обозначается общий тренд и предпосылки его изменения.

Инструменты и методы. В данном исследовании были использованы библиотеки `python`, а также инструмент `Stocker` [6] из открытой библиотеки, среды разработки `Visual Studio` и `Anaconda3`. Для обеспечения совместимости библиотек и повышения качества моделей был проведен ряд изменений в исходных кодах инструментов. В результате был получен единый вычислительный комплекс прогнозирования стоимости финансового инструмента.

Первоначально нами были импортированы исторические данные по котировкам акций `Caterpillar (CAT)` за период 02.01.1962 – 27.03.2018 (Рис. 1). Данные расположены в открытом доступе и импортированы с помощью инструмента `quandl` [3].

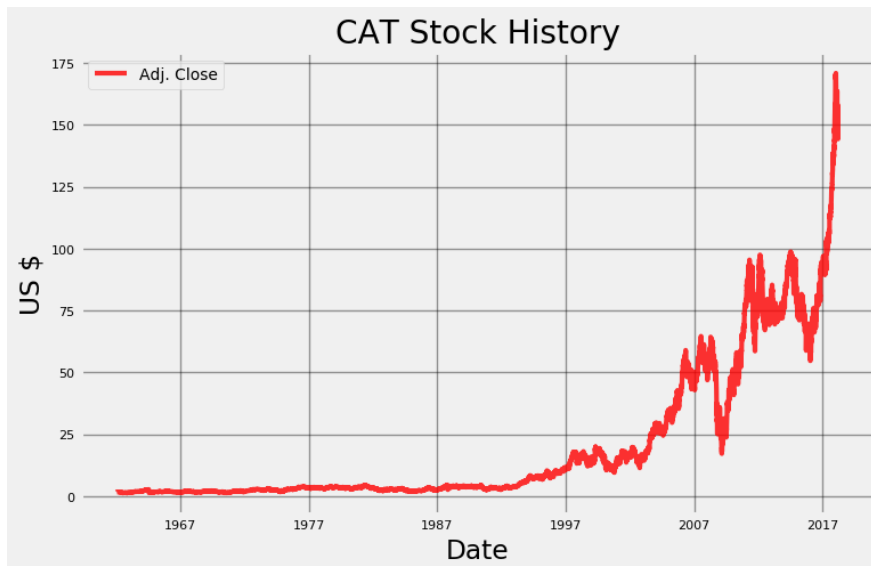


Рис. 1 – Тренд курса акций CAT за 1962-2018 гг.

Также для визуального анализа построены дневное изменение цены и дневная капитализация в процентах относительно среднего значения (рис.2). Как видно, с увеличением объема данных заметными становятся определенные зависимости и повторные флуктуации через неравные промежутки времени.

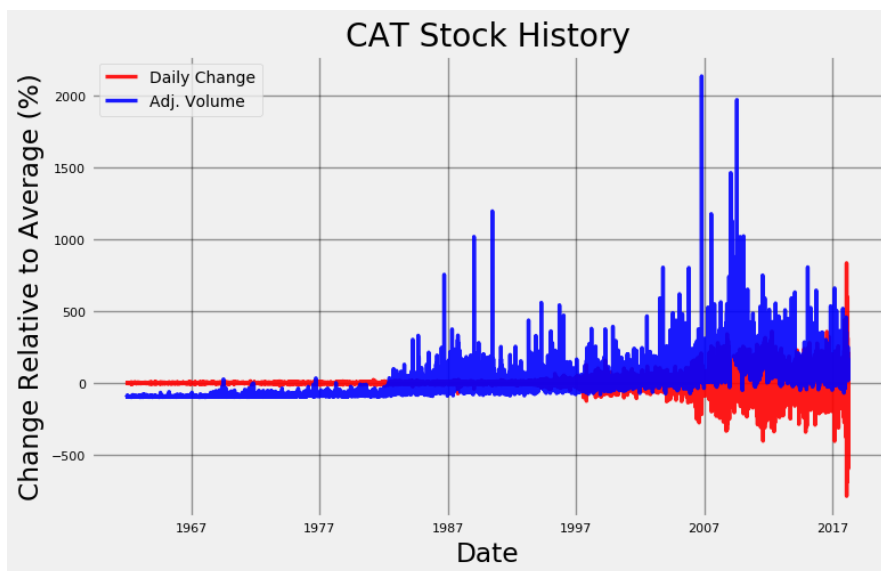


Рис. 2 – дневное изменение цены и дневная капитализация относительно среднего значения за 1962-2018 гг., %

Предсказания осуществляются с помощью аддитивной модели, которая рассматривает временные ряды как комбинацию тренда и сезонных изменений в разных временных масштабах (Рис. 3) [4][5]. Цена, предсказанная с помощью модели на 25.06.2018 составила \$176,90.

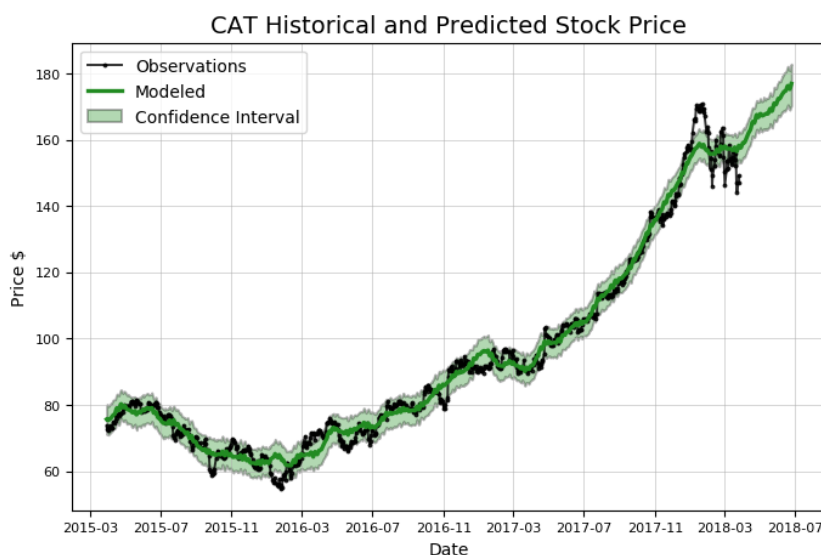


Рис. 3 – График статистических и прогнозных значений курса акций САТ

Для оценки качества модели программа выводит следующую статистику (рис. 4):

- предсказанная цена на 24.03.2018 – \$138,91;
- действительная цена на 23.03.2018 – \$144,29;
- средняя абсолютная ошибка на обучающем наборе данных – \$2.09;
- средняя абсолютная ошибка на тестовом наборе данных – \$12.32;
- когда модель предсказывала рост, цена росла в 58,05% случаев;
- когда модель предсказывала падение, цена падала в 42.67% случаев;
- реальные значения находились в 80% доверительном интервале в 29,60% случаев.

На основе выведенной статистики можно сделать вывод о довольно низком качестве модели.

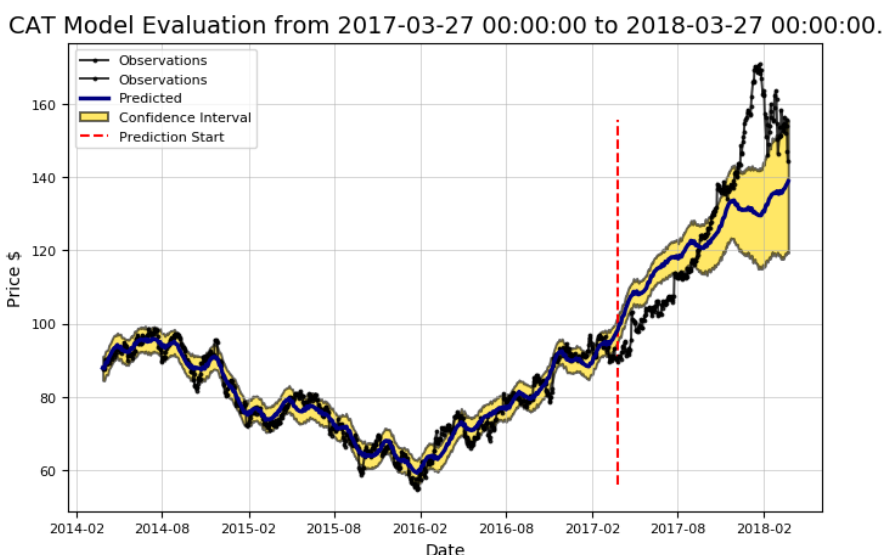


Рис. 4 – Оценка качества прогноза курса акций CAT за 27.03.2017 – 27.03.2018

Для улучшения качества модели был проведен дополнительный анализ статистики. Первоначально анализ строился на задании коэффициента масштаба распределения весов «контрольным точкам» (критическим точкам), в которых тренд графика меняет свое направление [6]. На рис. 5 показана визуальная оценка количества и положения данных точек на общем графике тренда с помощью построения графика второй производной.

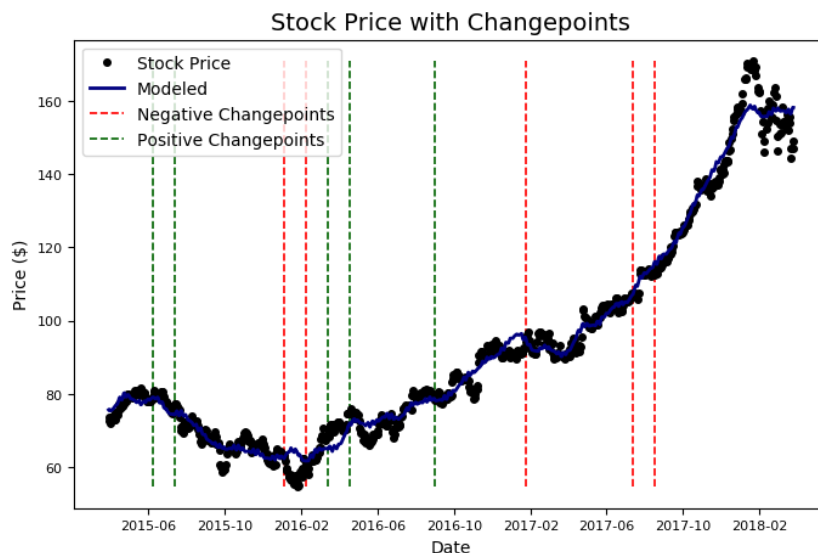


Рис. 5 – График курса акций САТ и контрольные точки

Изменяя веса добиваемся наиболее оптимального обучения модели (рис.

б)

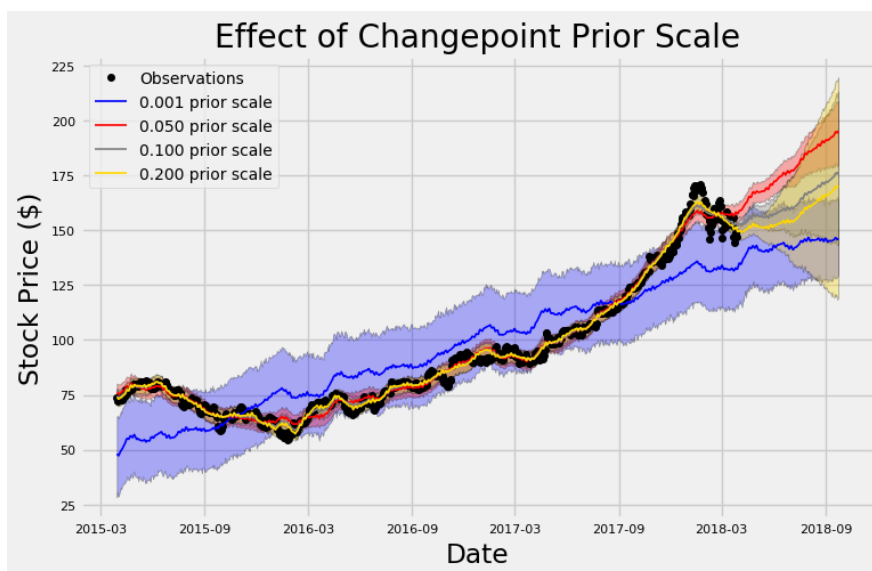


Рис. 6 – Эффект изменения весов контрольных точек

Таким образом, нам удалось добиться того, что реальные значения находились в 80% доверительном интервале с вероятностью 80.40% (рис. 7), что указывает на значительное повышение качества модели. С другой стороны, стоит отметить, что средняя ошибка выросла, что указывает на необходимость

выбора между «сужением» доверительного интервала и повышением вероятности попадания в него [2].

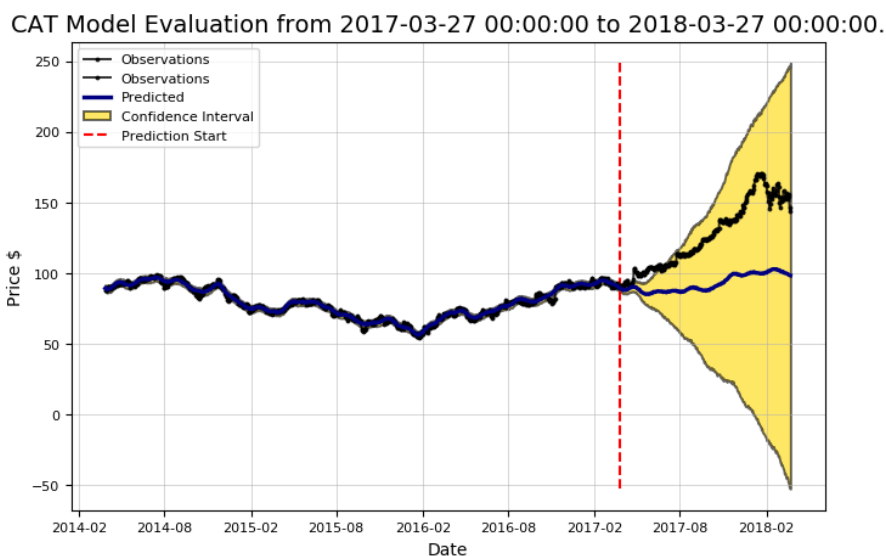


Рис. 7 – Оценка качества прогнозной модели за 27.03.2017 – 27.03.2018

Увеличим объем выборки для обучения расширив ее до диапазона 04.01.2012 – 03.01.2017. Подбирая необходимые веса получаем следующую статистику: действительные значения попадали в 80% доверительный интервал с вероятностью 94,80% (рис. 8)

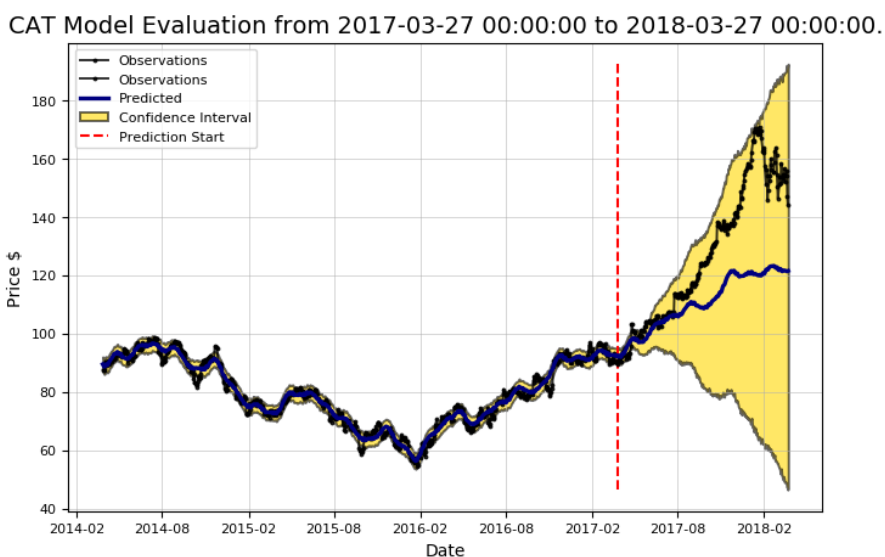


Рис. 8 – Оценка качества прогнозной модели за 27.03.2017 – 27.03.2018

Прогнозируя будущие значения котировок акций получаем следующие выходные данные для 2 и 5 -летнего интервала данных для обучения. Притом, прогнозирование делалось на следующие 30 (2-х летний период – рис. 9, 5-и летний период – рис. 10) и 365 (рис. 11, рис. 12 соответственно для 2 и 3-летнего периода) дней. Как видно, модели, обучение которых было основано на big data показывают большую точность предсказаний с меньшим доверительным интервалом.

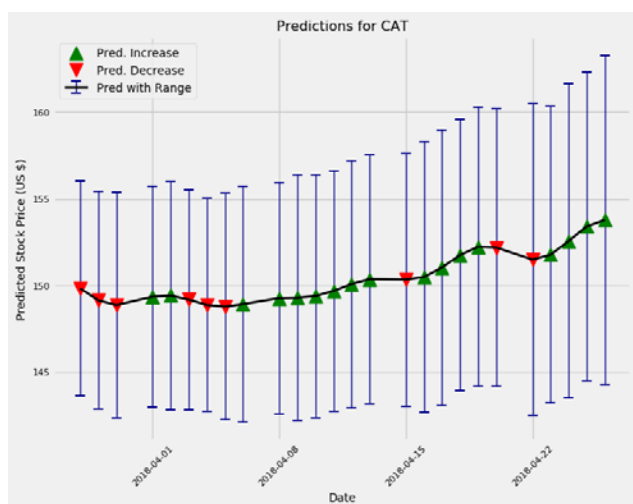


Рис. 9 – Прогнозные значения курса акций на 30 дней на основе данных за 2 года

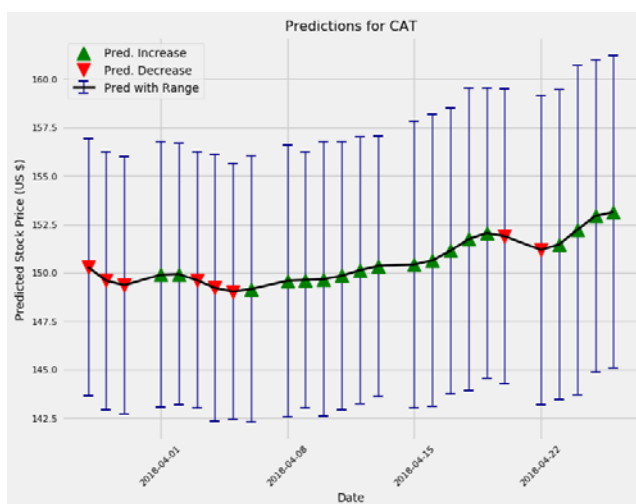


Рис. 10 – Прогнозные значения курса акций на 30 дней на основе данных за 5 лет

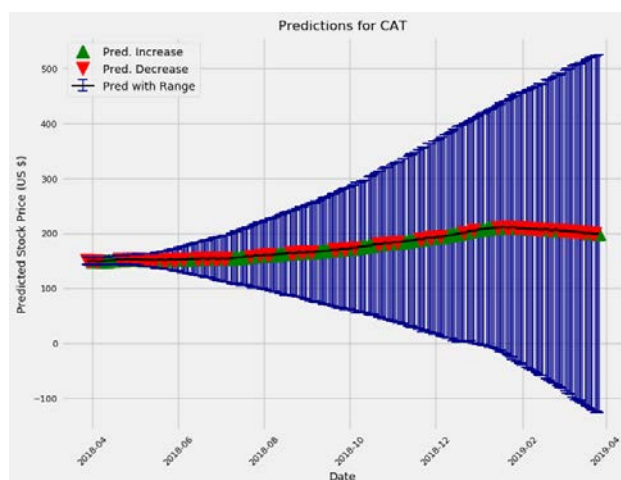
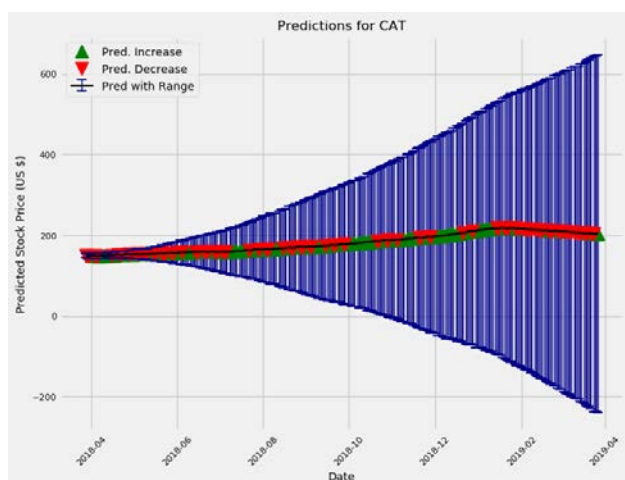


Рис. 11– Прогнозные значения курса акций на 365 дней на основе данных за 2 года

Рис. 12 – Прогнозные значения курса акций на 365 дней на основе данных за 5 лет

Обсуждения и дискуссии. На основе проведенного анализа можно сделать вывод о все возрастающем значении big data и необходимости использования специальных инструментов для их сбора, обработки и анализа. Актуальность рассматриваемого вопроса определяется ростом неопределенности на финансовых рынках и необходимостью задания доверительного интервала прогноза с определенной точностью и достоверностью для снижения возможных рисков внезапных флуктуаций котировок интересующего нас финансового инструмента. Результат прогнозирования может быть полезен инвесторам, желающим разработать торговую стратегию, а также стэйкхолдерам компании для определения ее позиции на рынке.

Библиографический список

1. Кораблев, А.В. Оптимизация информационных потоков в системе управления банковской деятельностью / А.В. Кораблев. – М.: ООО «Актуальность.РФ», 2017 – 136 с.
2. Кораблев, А.В., Петрушова, М.В. Информационное обеспечение системы управления современным предприятием / А.В. Кораблев, М.В. Петрушова // Проблемы развития предприятий: теория и практика. 2018. № 3. С. 35-39.
3. Официальный сайт библиотеки Quandl [Электронный ресурс] / «Quandl». – Режим доступа: <https://www.quandl.com> (дата обращения: 15.04.2019).
4. Прогнозирование фондового рынка на Python с помощью Stocker [Электронный ресурс] / Режим доступа: <https://neurohive.io/ru/tutorial/prognozirovanie-rynka-python-stocker/> (дата обращения 14.04.2019)
5. Hans-Georg Muller, Fang Yao. Functional Additive Models, July 2008
6. Stocker: A Stock Analysis and Prediction Toolkit using Additive Models [Электронный ресурс] / Режим доступа: <https://github.com/WillKoehrsen/Data-Analysis/tree/master/stocker>

Оригинальность 99%