Интернет-журнал «Отходы и ресурсы» https://resources.today Russian Journal of Resources, Conservation and Recycling

2020, №4, Том 7 / 2020, No 4, Vol 7 https://resources.today/issue-4-2020.html

URL статьи: https://resources.today/PDF/14INOR420.pdf

DOI: 10.15862/14INOR420 (http://dx.doi.org/10.15862/14INOR420)

Ссылка для цитирования этой статьи:

Рудзейт О.Ю., Зайнетдинов А.Р., Недяк А.В., Рагулин П.Г. Прогнозирование цены акции с помощью метода регрессионного анализа // Интернет-журнал «Отходы и ресурсы», 2020 №4, https://resources.today/PDF/14INOR420.pdf (доступ свободный). Загл. с экрана. Яз. рус., англ. DOI: 10.15862/14INOR420

For citation:

Rudzeyt O.Ui., Zainetdinov A.R., Nedyak A.V., Ragulin P.G. (2020). Predicting the price of a stock using regression analysis. *Russian Journal of Resources, Conservation and Recycling*, [online] 4(7). Available at: https://resources.today/PDF/14INOR420.pdf (in Russian) DOI: 10.15862/14INOR420

УДК 004.021

ГРНТИ 50.41.25

Рудзейт Олег Юрьевич

ФГОУ ВО «Дальневосточный федеральный университет», Владивосток, Россия Магистрант E-mail: rudzeyt18@mail.ru

Зайнетдинов Артём Ришатович

ФГОУ ВО «Дальневосточный федеральный университет», Владивосток, Россия Магистрант E-mail: zainet.ar@gmail.com

Недяк Антон Витальевич

ФГОУ ВО «Дальневосточный федеральный университет», Владивосток, Россия Магистрант E-mail: nedyaq@gmail.com

Рагулин Петр Григорьевич

ФГОУ ВО «Дальневосточный федеральный университет», Владивосток, Россия Профессор кафедры «Компьютерных систем» Кандидат технических наук E-mail: ragulin_p@mail.ru

Прогнозирование цены акции с помощью метода регрессионного анализа

Аннотация. На данный момент отмечается высокая тенденция к вложению финансовых средств для дополнительного заработка на биржах акций, их спекуляциях и т. д. На торговых платформах работает большое количество людей, покупает или продает акции, от чего и меняется их цена в различные периоды. На основе их цены формируются графики, по которых производят технический и фундаментальный анализ для построения стратегии заработка и прогнозирования даже не цены акции, а направления движения ее стоимости в сторону роста или уменьшения. Для получения ответа на данный вопрос прибегают к различным методам изучения исторических данных акций, изучению их качественных показателей, а также математическим методам прогноза.

В данной статье рассматривается пример работы алгоритма прогнозирования цены акции на момент закрытия торгового дня на рынке.

Метод, который позволит получить информацию о цене акции, основывается на использовании линейного регрессионного анализа. Данный метод лучше всего подходит для исследования, т. к. графики цен акций имеют линейный вид. Метод регрессионного анализа позволяет учитывать такие факторы, как цикличность повторения трендов и тенденций увеличения или уменьшения стоимости акции, длина периода торгов, в течении которого происходит получение исторических данных и их исследование и др. Также авторами была изучена работа алгоритма прогнозирования на основе регрессионного анализа и изучены факторы, влияющие на точность — размер обучающей выборки, количество исторических данных, количество дней, для которых будет спрогнозирована цена в будущем.

В ходе исследования была получена информация о точности прогнозирования цен акции, которая составляет 96 % в зависимости от валюты, в которой торгуется акция и компании, которая ее выпустила. Стоит отметить, что на точность прогноза также влияет объем исторических данных и финансовое положение корпорации на рынке.

Материалы представляют практическую значимость для развития систем поддержки принятия решений в экономических отраслях. Разработанную модель в дальнейшем можно использовать как основу для помощи в принятии решений для игры на бирже.

Ключевые слова: прогноз; цена; акция; анализ; алгоритм; выборка; объем

Введение

В техническом анализе имеется достаточно много способов оценки стоимости акции и определения тренда на повышение или понижение ее цены для этого используются такие показатели как отношение текущей рыночной стоимости акции к чистой прибыли компании на 1 акцию, отношение чистой прибыли компании к количеству размещенных акций на бирже, отношение дивидендов по обыкновенным акциям к чистой прибыли на одну акцию и т. п. [1–2]. Изучение представленных показателей относятся к одному из много численных способов получения информации о будущей стоимости акции, которые используют аналитики.

Для получения какой-либо информации о тренде движения стоимости акции также стоит использовать математические методы исследования рынка: регрессионный анализ, кластерный анализ, метод опорных графов.

В случае применения методов машинного обучения для обработки торговых данных, наиболее часто применяют метод технического анализа, цель которого заключается в том, чтобы понять, может ли алгоритм точно определять паттерны поведения акции во времени. Тем не менее, методы машинного обучения могут быть применены также для оценки и прогнозирования результатов деятельности компании для дальнейшего их использования при фундаментальном анализе [3–4].

Одним из методов был выбран метод линейной регрессии, который устанавливает зависимость одной переменной от другой или нескольких других переменных с линейной функцией зависимости также исходя из статистических данных данный метод показывает одну из наиболее точных оценок прогнозирования и наименьшими показателями функции потерь.

Линейная регрессия имеет привлекательную модель, потому что ее представление является достаточно простым для пользователя, который не знаком с математическими методами анализа данных.

Представление – это линейное уравнение, объединяющее определенный набор входных значений (X) решений, к которому является прогнозируемый вывод для этого набора входных значений (y) [5–6].

Примером наиболее простой задачи линейной регрессии является подача на вход значений одной переменной X, а на выходе будет вычислено значение одной переменной Y. Подобная модель представляется в виде:

$$Y = B0 + B1 * X$$

В случае многомерного анализа, линия превращается в плоскостью или гиперплоскости [7].

Материалы и методы

Регрессионный анализ — это статистический аналитический метод, позволяющий вычислить предполагаемые отношения между зависимой переменной одной или несколькими другими переменными. Применяя регрессионный анализ, можно моделировать отношения между представленными переменными, а также предсказываемыми значениями на основе определенной математической модели. На сегодня существует множество видов регрессий: бинарная регрессия, нелинейная регрессия, мультиномиальная логистическая регрессия и т. д. Для предсказывания цены акции выберем метод линейной регрессии.

В математическом анализе линейная регрессия является методом аппроксимации зависимости между входными и выходными данными, используя линейную математическую модель.

В регрессионном анализе входные данные также именуются предикторными переменными (регрессорами), а выходные данные – критериальными.

При рассмотрении зависимости между одной входными и выходными данными в случае линейной регрессии зависимость определяется уравнением: Y = ax + b и строится соответствующая этому тождеству прямая, называемая линией регрессии (рисунок 1).

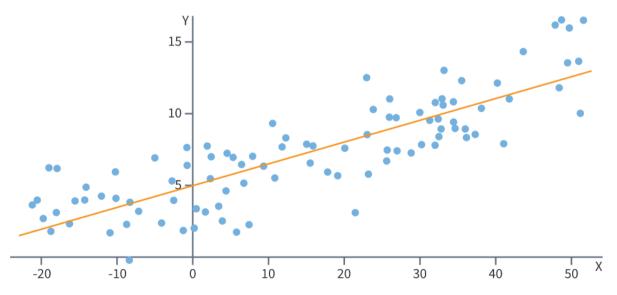


Рисунок 1. Пример линейной зависимости

Figure 1. Linear dependence example

Коэффициенты а и b, называемые также параметрами модели, определяются так, чтобы сумма квадратов отклонений точек, которые соответствуют реальным данным, от линии регрессии стремилась к минимуму. Поэтому коэффициенты чаще всего оцениваются методом наименьших квадратов.

Если имеется зависимость между несколькими входными данными и одной выходной переменными, то имеет место множественная линейная регрессия, уравнение которой имеет вид:

```
Y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n, где n - число входных переменных.
```

Очевидно, что в данном случае модель будет описываться гиперплоскостью.

Линейная регрессия была первым видом регрессионного анализа, который был изучен и обширно применен в различного рода прикладных задачах. Факторов быстрого развития данного метода стало то, что в линейных моделях оценивание параметров является наиболее простым, чем в других методах.

Линейная регрессия имеет множество практических применений. Большинство приложений попадают в одну из двух широких категорий:

- Если целью является прогнозирование, линейную регрессию можно использовать для подгонки модели к наблюдаемому набору данных.
- Если цель заключается в том, чтобы объяснить изменчивость выходной переменной, можно применить линейный регрессионный анализ для количественной оценки силы взаимосвязи между выходной и входными переменными¹.

Далее проведем несколько экспериментов с целью выяснения точности прогнозируемых данных описанным методом регрессии.

Для прогноза ежедневной цены акций в момент закрытия торгового для, используя данные за предыдущие N дней (горизонт прогноза равен 2 дням) будут использованы тридцатилетние исторические данные на стоимости десяти различных акций американских и русских компаний в период с 1990-01-01 по нынешнее время. Данные будут загружены с помощью библиотеки yfinance [8], которая позволяет получать данные об акциях с торговой платформы и библиотеки Pandas [9], которая используется при работе с Большими данными и работы со специальными структурами данных — фреймами. Для проведения экспериментов был выбран язык программирования Python. Пример загруженных данных представлен на рисунке 2.

	Date	AAPL	WMT		TATNP.ME	LSRG.ME	TTLK.ME
0	1990-01-22	0.239298	3.398741		NaN	NaN	NaN
1	1990-01-23	0.242896	3.398741		NaN	NaN	NaN
2	1990-01-24	0.244696	3.418791		NaN	NaN	NaN
3	1990-01-25	0.245595	3.358640		NaN	NaN	NaN
4	1990-01-26	0.235700	3.378692		NaN	NaN	NaN
7789	2020-12-18	126.660004	145.949997		475.799988	850.200012	0.4220
7790	2020-12-21	128.229996	145.970001		462.399994	854.200012	0.4120
7791	2020-12-22	131.880005	144.199997		467.399994	868.599976	0.4200
7792	2020-12-23	130.960007	143.220001		469.500000	859.400024	0.4235
7793	2020-12-24	131.970001	143.500000		468.600006	854.000000	0.4190

Рисунок 2. Пример выгруженных данных

Figure 2. Example of uploaded data

¹ https://wiki.loginom.ru/articles/linear-regression.html – Линейная регрессия (Linear Regression).

На следующем этапе требуется подготовить представленные данные для исследования методом регрессионного анализа, а именно, удалить колонку с датами и разделить показания цен акции на различные фреймы для дальнейшего изучения алгоритма.

Для изучения точности прогнозирования методом линейной регрессии было произведено сравнение графиков истинной цены, взятой за последние 30 лет и цены акции, построенной с помощью метода регрессионного анализа на примере акции компании Apple (рисунок 3). Для получения оптимальной модели с высокой точностью исторические данные будут разделены на 2 части: 80 % данных будет использовано для обучения модели, а 20 % данных будет применено для тестирования модели и расчета точности показаний. Для удобства представления данных было принято решение представить цену и прогноз акций компании Apple за предыдущие 3 месяца.

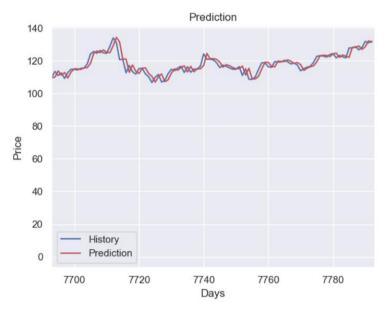


Рисунок 3. Графики цен акции компании Apple

Figure 3. Apple stock price charts

Исходя из представленной математической модели стоимости акции Apple можно сказать о том, что линейная регрессия почти полностью повторяет движение тренда стоимости и имеет точность 96 %. В таблице 1 перечислены показания точности моделей при различных размерах обучающих и тестовых выборках, полученных авторами в ходе анализа модели на примере акций Apple. Также было установлено, что точность прогнозирования может варьироваться в зависимости от акций различных компаний.

Таблица 1 / Table 1
Точность прогноза в зависимости от размера обучающей выборки
Prediction accuracy depending on the size of the training sample

Обучающая выборка (%)	Тестовая выборка (%)	Точность прогноза (%)
50	50	76,2
55	45	83,4
60	40	87,1
65	35	92,88
70	30	93,9
75	25	94,8
80	20	95,3
85	15	95,9

Для оценки точности алгоритма на различном объеме выборки обучающих данных проведем эксперимент при изменении периода выборки исторических данных и сократим период времени, за который была получена информация с 30 лет до 1 года и сравним полученные результаты с результатами предыдущего эксперимента при сохранении доли обучающей выборки от общего количества данных (рисунок 4).

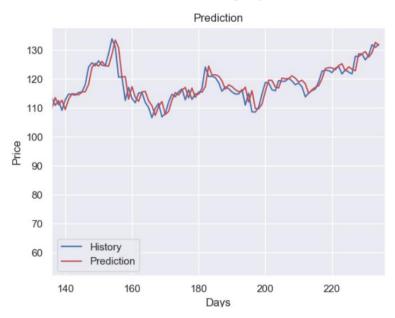


Рисунок 4. Сравнение цен акции компании Apple

Figure 4. Apple stock price comparison

В данном случае на обучающей выборке, взятой за 1 год, наблюдается наиболее сильное расхождение графиков, чем в предыдущем в примере. Также точность полученных данных, построенных методом линейной регрессии, упала с 96 % до 91 %. Это связано с тем, что за 30 лет было получено почти 8000 значений для обучения, а за 1 год лишь 250. Также возможны появления аномалий в работе алгоритма по причине того, что метод линейной регрессии не учтет аномалии, например, во время финансового кризиса или моментов, когда компания приходила в упадок, также рост стоимости акций компании во время презентации новых устройств, либо поглощения какой-либо компании исследуемой компанией в случае рассмотрения примера с Microsoft.

Для проверки точности алгоритма различных акций российских и зарубежных компаний были взяты такие компании как Apple, Яндекс, Bank of America, Татнефть и др. Прогнозирование цены производилось на момент 27 декабря 2020 года с историческими данными за последние 30 лет, обучающая выборка составила 70 % от полученных данных, а тестовая выборка составила 30 %. Прогноз цены для сравнения был взят также на 27 декабря (таблица 2).

Таблица 2 / Table 2 Сравнение реальной цены акции и спрогнозированной Comparison of the real share price and the forecast price

Тикер Акции	Стоимость акции, долл.	Спрогнозированная стоимость, долл.
APPL	131,97	131,9
WMT	143,5	143,39
IBM	124,69	123,66
MU	70,58	70,32
BA	217,15	219,038

Тикер Акции	Стоимость акции, долл.	Спрогнозированная стоимость, долл.
AXP	117,35	116,89
YNDX	68,49389417	69,1541791
TATNP	6,369063772	6,373242877
LSRG	11,69335142	11,67689281

Для наглядности представим полученные данные в виде гистограммы (рисунок 5).

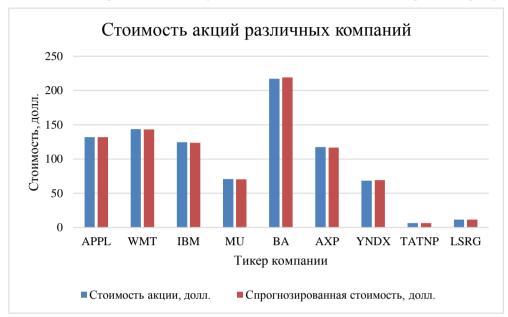


Рисунок 5. Сравнение реальной стоимости и спрогнозированной стоимости

Figure 5. Comparison of real cost and projected cost

Исходя и построенной диаграммы можно отметить достаточно высокую точность алгоритма. Расхождение в цене не превышают 2 долларов для компаний с большой стоимостью акций и не более 1 доллара для Российских компаний. В данном случае на стоимость закрытия акции будет также влиять и валюта страны, где находится российская компания [10]. Обычно, за день у российских компаний наблюдаются более динамичные скачки, чем у американских или европейских компаний. Это объясняется более слабым курсом валюты, чем у развитых стран.

Заключение

Исследование построения системы поддержки принятия решений на основе одного из методов машинного обучения, а именного построения линейной регрессии позволяет помочь провести технический анализ для определенной компании и принять решение о покупке или продаже акции.

Использование навыков математического анализа и собственного опыта позволит пользователю строить более точные прогнозы при кратковременных сделках и выстроить оптимальную стратегию на торговый день. Модель представленную в рамках данной статьи можно скорректировать, изменяя выборку исторических данных, размер обучающей и тестируемой выборок можно добиться оптимальной точности прогноза акции. Для применения других стратегий можно получать информацию, например, цены не на момент закрытия торгового дня, а цены акции каждый час в период торгов.

В рамках данной статьи можно также сделать следующие выводы:

- Линейная регрессия имеет достаточно высокую точность прогноза 96 % для того, чтобы принять финансовое решение.
- Точность прогноза падает с уменьшением количества исторических данных.
- Точность прогноза может варьироваться в зависимости от компании, т. к. у различных компаний своя история развития, а значит, и различные исторические данные.
- На прогноз также влияет валюта страны, в которой располагается головной офис компании, представленной на рынке.

ЛИТЕРАТУРА

- 1. Ивасенко А.Г. Рынок ценных бумаг: инструменты и механизмы функционирования. Москва: КНОРУС, 2017. 272 с.
- 2. Gottschlich, J., & Hinz, O. (2014). A decision support system for stock investment recommendations using collective wisdom. Decision Support Systems, 59, 52–62. http://dx.doi.org/10.1016/j.dss.2013.10.005.
- 3. Васильева, Л.С. Финансовый анализ / Л.С. Васильева, М.В. Петровская. М.: КноРус, 2017. 880 с.
- 4. Mohamed Masry (2017). The Impact of Technical Analysis on Stock Returns in an Emerging Capital Markets (ECM's) Country: Theoretical and Empirical Study International Journal of Economics and Finance, Volume 9(3): 91. Doi: 10.5539/ijef.v9n3p91.
- 5. Пономарев И.В., Славский В.В. О геометрической интерпретации метода наименьших квадратов // Известия Алтайского гос. ун-та. -2012. -№ 1-1(73). -C. 119-121.
- 6. Маккинли, У. Python и анализ данных / У. Маккинли. М.: ДМК, 2015. 482 с.
- 7. О. Жерон. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow. Концепции, инструменты и техники для создания интеллектуальных систем / О. Жерон, М.: Вильямс, 2018, 688 с. ISBN 978-5-950-02962-2.
- 8. Плас, Джейк Вандер Python для сложных задач. Наука о данных и машинное обучение. Руководство / Плас Джейк Вандер. М.: Питер, 2018. 759 с.
- 9. Ажаханова, Д.С. Современные подходы к стратегическому управлению // Современные тенденции в экономике и управлении: новый взгляд. 2018. № 19. С. 16–18.
- 10. Астахов А.А., Гребеник В.В. Проблемы и перспективы развития рынка ценных бумаг как инструмента инвестирования в России // Вестник Евразийской науки, 2018 №3, https://esj.today/PDF/52ECVN318.pdf (доступ свободный).

Rudzeyt Oleg Uirievich

Far Eastern federal university, Vladivostok, Russia E-mail: rudzeyt18@mail.ru

Zainetdinov Artem Rishatovich

Far Eastern federal university, Vladivostok, Russia E-mail: zainet.ar@gmail.com

Nedvak Anton Vitalievich

Far Eastern federal university, Vladivostok, Russia E-mail: nedyaq@gmail.com

Ragulin Petr Grigorievich

Far Eastern federal university, Vladivostok, Russia E-mail: ragulin_p@mail.ru

Predicting the price of a stock using regression analysis

Abstract. At the moment, there is a high tendency towards investing financial resources for additional earnings on stock exchanges, their speculations, etc. A large number of people work on trading platforms, buy or sell stocks, which causes their price to change in different periods. On the basis of their prices, charts are formed, according to which technical and fundamental analysis is performed to build a strategy for earning and forecast not even the price of a stock, but the direction of movement of its value in the direction of growth or decrease. To obtain an answer to this question, they resort to various methods of studying the historical data of stocks, studying their qualitative indicators, as well as mathematical forecasting methods.

This article examines an example of how the algorithm for predicting the stock price at the close of the trading day on the market.

The method, which will provide information about the stock price, is based on the use of linear regression analysis. This method is best suited for research, because stock price charts are linear. The method of regression analysis allows you to take into account such factors as the cyclical recurrence of trends and tendencies of increase or decrease in the value of a stock, the length of the trading period, during which historical data is obtained and studied, etc. The authors also studied the work of the forecasting algorithm based on regression analysis and studied factors affecting accuracy – the size of the training sample, the amount of historical data, the number of days for which the price will be predicted in the future.

In the course of the study, information was obtained on the accuracy of forecasting stock prices, which is 96 % depending on the currency in which the stock is traded and the company that issued it. It should be noted that the accuracy of the forecast is also affected by the volume of historical data and the financial position of the corporation in the market.

The materials are of practical importance for the development of decision support systems in economic sectors. The developed model can be used as a basis for helping to make decisions for trading on the stock exchange.

Keywords: forecast; price; share; analysis; algorithm; sample; size