**Министерство науки и высшего образования Российской Федерации**

**Федеральное государственное автономное образовательное учреждение высшего образования**

**«КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

ИНСТИТУТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

КАФЕДРА ТЕХНОЛОГИЙ ПРОГРАММИРОВАНИЯ

Направление: 09.03.03 – Прикладная информатика

КУРСОВАЯ РАБОТА

**Прогнозирование финансового рынка на основе нейросетевых технологий на примере работы акций на золото.**

Студент 3 курса

Группы 09-753

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_ 2021 г.    \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Плотницкий А.С.

Научный руководитель

к.ф. – м.н.,

доцент

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_ 2021 г.    \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Еникеева З.А.

Казань – 2021

Содержание

[Введение 3](#_Toc42424608)

[1. Проектирование информационной системы](#_Toc42424609) 5

[1.1. Предназначение системы 5](#_Toc42424610)

[1.2. Выбор программных средств реализации 5](#_Toc42424611)

[1.3. Конфигурация базы данных](#_Toc42424615) 8

[2. Обзор методов машинного обучения, используемых в работе](#_Toc42424616) 10

[2.1. CNN](#_Toc42424616) 10

[2.2. LSTM](#_Toc42424616) 10

[3. Реализация программы для классификации временных рядов финансового рынка 1](#_Toc42424619)3

[3.1. Подготовка данных 1](#_Toc42424620)3

[3.2. Конфигурирование и проектирование нейронной сети 1](#_Toc42424621)6

[3.3. Тестирование программы](#_Toc42424624) 18

3.[4. Тестирование программы на других временных периодах 2](#_Toc42424628)3

[Заключение](#_Toc42424630) 28

[Список использованных источников](#_Toc42424631) 29

### 

### Введение

Финансовый рынок на сегодняшний день является быстро развивающейся сферой, которая вполне может стать одним из основополагающим фактором экономики в целом. Внимание миллионов людей приковано к изменениям которые происходят в процессе торгов, каждый из которых хочет заработать на этих самых изменениях, но на финансовый рынок действует огромное количество различных факторов и без какой либо подготовки практически невозможно предугадать дальнейшие изменения. С этой проблемой люди, - за десятки лет, с учетом множества ситуаций происходящих в мире. С помощью данных алгоритмов люди могут принимать решения в реальном времени. Хоть алгоритм и не может охватить все факторы от которых зависит какая либо акция, так как стоимость в данном случае во многом зависит от человеческих факторов (политика которую выбрали лидеры стран, конфликты крупных компаний и даже репутация отдельно стоящих людей), но все равно точность предсказания зачастую очень близка к реальным значениям. Так же не стоит забывать о том, что машинное обучение так же очень быстро развивающаяся сфера которая ни на секунду не стоит на месте, работа в оптимизации отдельных алгоритмов, изучение новых подходов, все это будет с течением времени только увеличивать точность в вопросах предсказывания динамики финансового рынка.

Цель работы: исследование предметной области и создание программы, позволяющей производить предсказания стоимости золота.

В ходе работы было выявлено несколько основных задач:

1. Изучение основных принципов устройста финансового рынка.

2. Составление стека технологий, с помощью которого будет реализована программа.

3. Создание базы данных.

4. Проектирование архитектуры нейронной сети.

5. Обучение и тестирование полученной модели 1.

6. Обучение и тестирование полученной модели 2.

7. Краткосрочное и среднесрочное прогнозирование.

8. Сравнение и анализ результатов.

### Проектирование информационной системы

### . Предназначение системы

Конечный продукт служит для прогнозирования и классификации временного ряда на основе пользовательских данных, а так же для произведения предсказаний длительностью до 9 дней. Представляет собой инструмент для анализа финансового рынка отображающее динамику изменения любой акции, в данном конкретном случае, золота.

### 1.2. Выбор программных средств реализации

Jupyter Notebook – Браузерная IDE позволяющая запускать код модулями и позволяет сохранять выходные данные каждого модуля. Все в одном месте: Jupyter Notebook - это интерактивная веб-среда, которая объединяет в одном документе код, форматированный текст, изображения, видео, анимацию, математические уравнения, графики, карты, интерактивные фигуры и виджеты, а также графические пользовательские интерфейсы.

Легко конвертировать: Jupyter поставляется со специальным инструментом nbconvert, который конвертирует записные книжки в другие форматы, такие как HTML и PDF. Другой онлайн-инструмент, nbviewer, позволяет нам отображать общедоступную записную книжку прямо в браузере.

Независимость от языка: архитектура Jupyter не зависит от языка.

Разделение между клиентом и ядром позволяет писать ядра на любом языке.

Легко создавать оболочки ядра: Jupyter предоставляет легкий интерфейс для языков ядра, которые можно обернуть в Python. Ядра оболочки могут реализовывать дополнительные методы, особенно для завершения кода и проверки кода.

Легко настроить: интерфейс Jupyter можно использовать для создания полностью настраиваемого интерфейса в Jupyter Notebook (или другом клиентском приложении, таком как консоль).

Расширения с настраиваемыми магическими командами: создавайте расширения IPython с настраиваемыми магическими командами, чтобы сделать интерактивные вычисления еще проще. Многие третьи-

существуют партийные расширения и магические команды, например, магия %% cython, которая позволяет писать код Cython прямо в записной книжке.

Воспроизводимые эксперименты без стресса: записные книжки Jupyter помогут вам с легкостью проводить эффективные и воспроизводимые интерактивные вычислительные эксперименты. Это позволяет вам вести подробный учет вашей работы. Также,

простота использования Jupyter Notebook означает, что вам не нужно беспокоиться о воспроизводимости; просто выполняйте всю свою интерактивную работу в записных книжках, ставьте их под контроль версий и регулярно фиксируйте. Не забудьте преобразовать код в независимые компоненты многократного использования.

Инструмент обучения: Jupyter Notebook - это не только инструмент для научных исследований и анализа данных, но и отличный инструмент для обучения. Примером может служить IPython Blocks - библиотека, которая позволяет вам или вашим ученикам создавать сетки из красочных блоков.

Интерактивный код и исследование данных: пакет ipywidgets предоставляет множество стандартных элементов управления пользовательского интерфейса для интерактивного изучения кода и данных.

PostgreSQL – свободная и очень востребованная в мире IT реляционная СУБД, основными особенностями которой является легкая расширяемость, поддержка БД любой размерности.

Преимущества:

PostgreSQL может запускать динамические веб-сайты и веб-приложения в качестве опции стека LAMP.

Ведение журнала с упреждающей записью в PostgreSQL делает его очень отказоустойчивой базой данных.

Исходный код PostgreSQL находится в свободном доступе по лицензии с открытым исходным кодом.

MatPlotLib – Библиотека которая позволяет визуализировать данные в виде графиков и т.д.

Преимущества:

Быстрый и эффективный

Кросплатформенность

Высокое качество графики

NumPy – Библиотека которая добавляет поддержка многомерных массивов и матриц, а также вносит большой вклад в вычисления.

Преимущества:

NumPy использует гораздо меньше памяти для хранения данных.

NumPy позволяет создавать n-мерные массивы.

Математические операции над n-мерными массивами NumPy.

Поиск элементов в массиве NumPy.

Pandas – Библиотека позволяющая работать с временными рядами.

Преимущества:

Обширный набор функций.

Эффективно обрабатывает большие данные.

Делает данные гибкими и настраиваемыми.

Keras – Компактная и очень мощная надстройка над TensorFlow которая позволяет работать с нейросетями.

Преимущества:

Удобное и быстрое развертывание.

Качественная документация и поддержка большого сообщества.

Модульность.

Предварительно обученные модели.

Поддержка нескольких графических процессоров.

### 1.3. Конфигурация базы данных

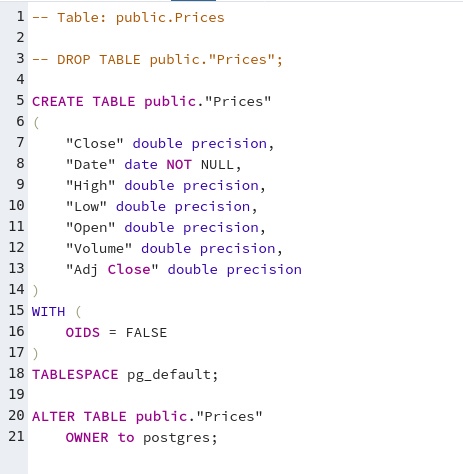


Рисунок 1.1 Скрипт создания таблицы базы данных

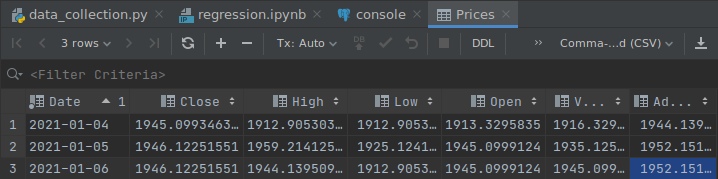


Рисунок 1.2 Заполнение таблицы

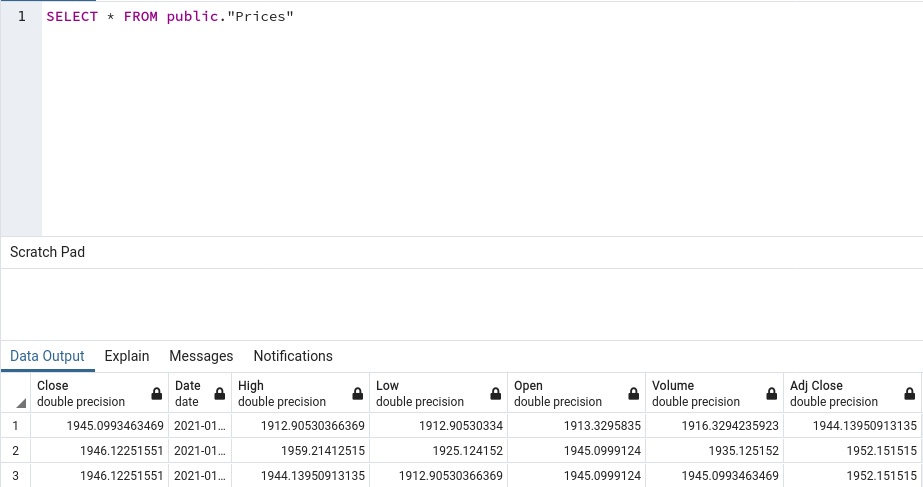


Рисунок 1.3 Проверка работоспособности базы данных

Теперь протестировав нашу базу данных мы можем быть уверены в том, что наша нейросеть сможет исправно получать данные. Для дальнейшего обучения и анализа данных на основе которых программа и будет делать свои предсказания

### 2. Обзор методов машинного обучения, используемых в работе

Машинное обучение это стремительно развивающаяся сфера, с каждым днем появляетс все больше новых архитектур для нейронных сетей, каждая из них имеет свое собственное предназначение, область задач с которыми они справляются лучше всего.

В ходе работы было принято решение использовать гибридную модель CNN-LSTM.

2.1 CNN

CNN - сетевая модель, предложенная Lecun et al. в 1998 г. CNN - это своего рода нейронная сеть с прямой связью, которая обладает хорошей производительностью при обработке изображений и обработке естественного языка. Его можно эффективно применять для прогнозирования временных рядов. Локальное восприятие и распределение веса CNN может значительно сократить количество параметров, тем самым повышая эффективность обучения модели. CNN в основном состоит из двух частей: сверточного слоя и объединяющего слоя. Каждый слой свертки содержит множество ядер свертки, и его формула вычисления показана в формуле. После операции свертки сверточного слоя характеристики данных извлекаются, но размеры извлеченных объектов очень велики, поэтому для решения этой проблемы и снижения затрат на обучение сети после слоя свертки добавляется слой объединения, чтобы уменьшить размерность объекта.

2.2 LSTM

LSTM - это сетевая модель, предложенная Шмидхубером и др. в 1997 г. LSTM - это сетевая модель, разработанная для решения давних проблем градиентного взрыва и исчезновения градиента в RNN. Он широко используется в распознавании речи, эмоциональный анализ и анализ текста, так как он имеет собственную память и может делать относительно точные прогнозы. В последние годы он стал применяться и в области прогнозирования фондового рынка. В стандартной RNN есть только один повторяющийся модуль, и его внутренняя структура проста. Тем не мение, четыре модуля LSTM аналогичны стандартным модулям RNN и работают в особом интерактивном режиме. Ячейка памяти LSTM состоит из трех частей: затвора забывания, входного затвора и выходного затвора. Данная модель включает в себя сверточные слои CNN для получения признаков из входных данных и запоминающие нейроны LSTM архитектуры которые позволяют поддерживать предсказания последовательности.

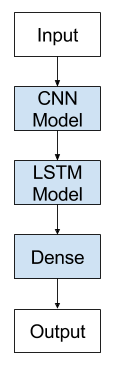


Рисунок 2.1 Архитектура нейронной сети

CNN-LSTM может обеспечить надежный прогноз цен на акции с высочайшей точностью прогнозов. Этот метод прогнозирования не только дает новую исследовательскую идею для прогнозирования цен на акции, но и дает практический опыт изучения данных финансовых временных рядов. Была выбрана именно эта архитектура в основном из за двух особенностей, а именно CNN имеет свойство выявлять особо очевидные параметры что очень помогает в проектировании признаков. А LSTM в свою очередь имеет свою собственную память что позволяет делать точные прогнозы.

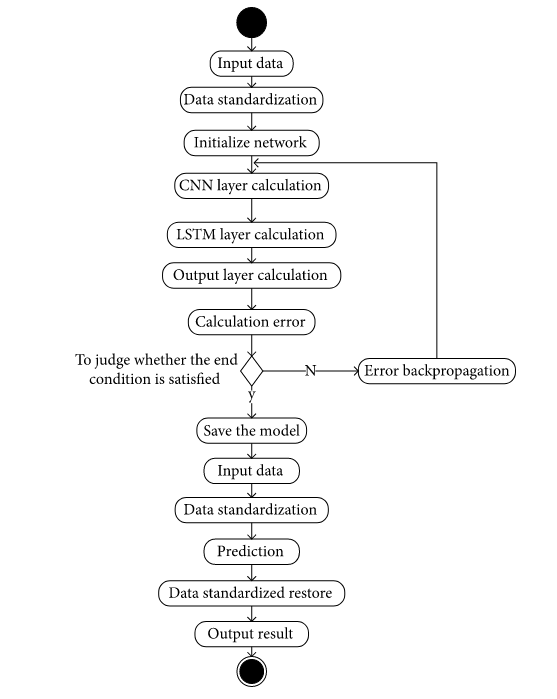


Рисунок 2.2 Расширенный пример CNN LSTM модели.

### 

### 3. Реализация программы для классификации временных рядов финансового рынка

Программа будет реализована в Jupyter Notebook, соответственно будет использован Python и достаточно обширное количество библиотек часть которых уже описана в п.1.2.

Библиотеки которые будут использованы в работе представленны на рисунке 3.1.

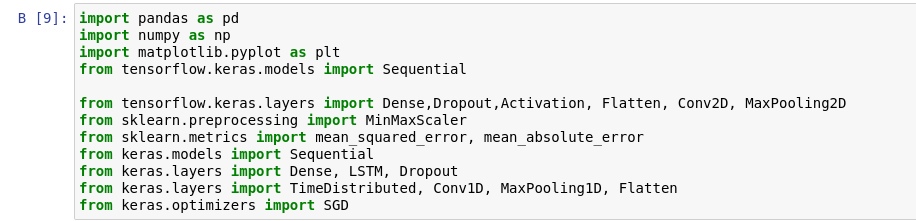


Рисунок 3.1 Использующиеся библиотеки.

Данные для работы нейронной сети будут получены путем парсинга с сайта Yahoo finance с помощью дополнительного скрипта в котором можно установить диапозон по датам получаемых данных, сделать это можно отдельно как для обучающей выборки так и для тестовых данных, независимо друг от друга.

### 3.1. Подготовка данных

Как и было сказано в предыдущем пункте данные мы получаем с помощью специального скрипта написанного также на языке Python.

Данные экспортированы с этого сайта (<https://finance.yahoo.com/quote/GC=F/>)

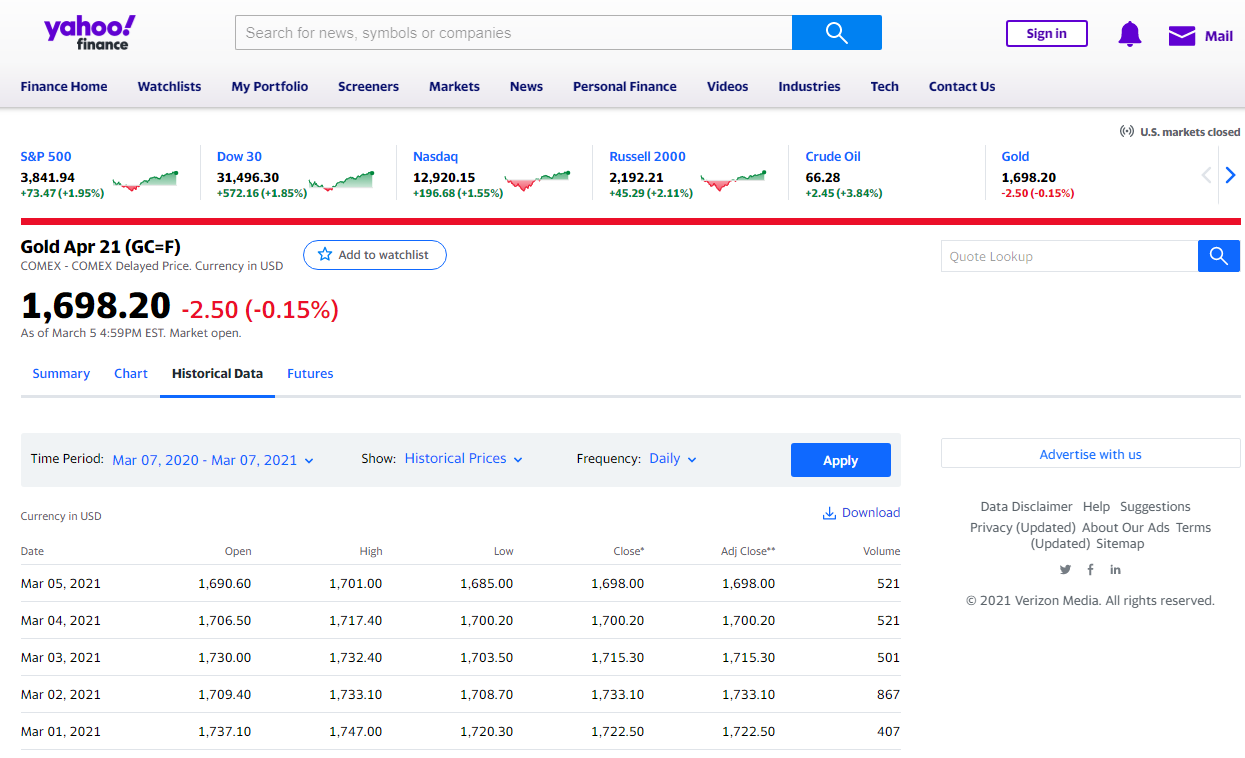


Рисунок 3.1.1 Страница акции с сайта Yahoo finance.

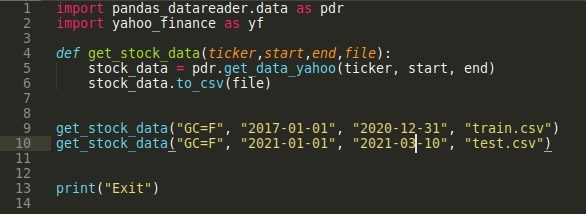


Рисунок 3.1.2 Скрипт с помощью которого мы получаем данные.

Как видно на рис.3.2 для обучающей выборки мы берем данные с 2017.01.01 по 2020.12.31. Для тестовой выборки мы берем данные с 2021.01.01 и по сегодняшний день.

Полученный датасет имеет следующий вид:

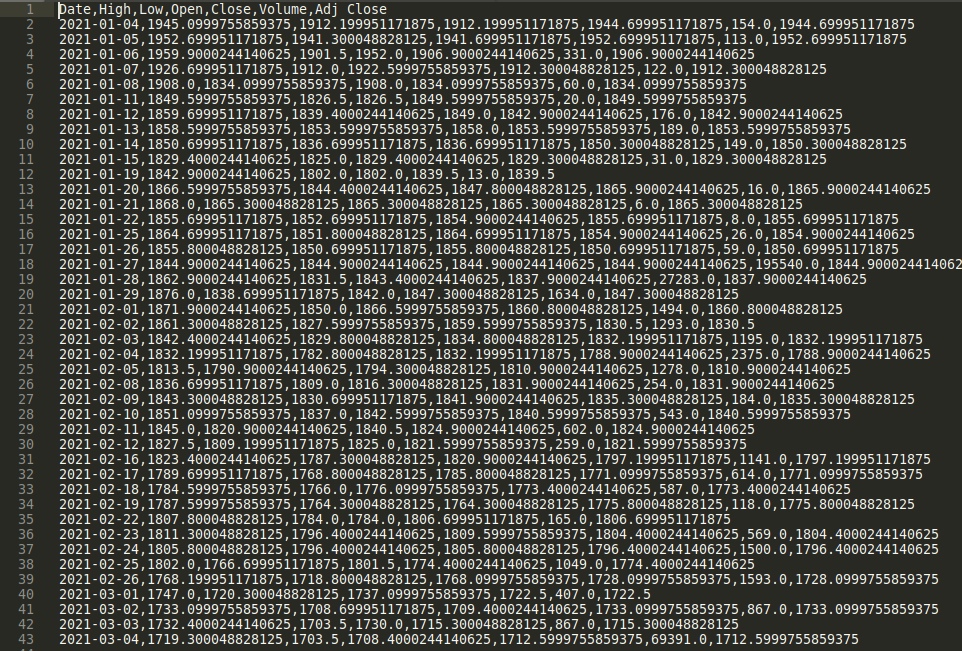




Рисунок 3.1.3 Датасет.

Получив данные с ними нужно поработать, а именно провести их нормализацию делается это для того, чтобы полученные результаты были более точными.



Рисунок 3.1.4 Фрагмент кода которые производит нормализацию.

### 3.2. Конфигурирование и проектирование нейронной сети

Теперь когда мы подготовили данные нужно выбрать функцию активации. В данной работе будет использована сама распространенная функция, а именно Relu. Она нормализирует данные в диапазоне [0,1] и добавляет нелинейность.

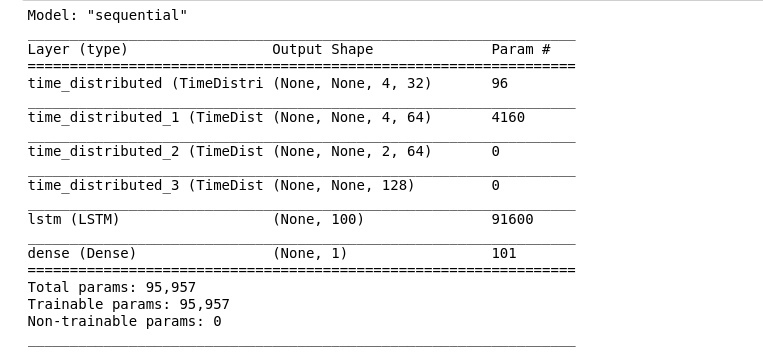


Рисунок 3.2.1 Архитектуры первой сети.

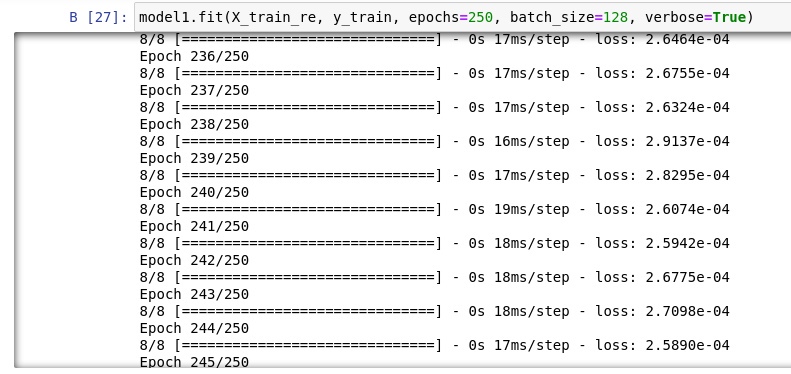


Рисунок 3.2.2 Процесс обучения

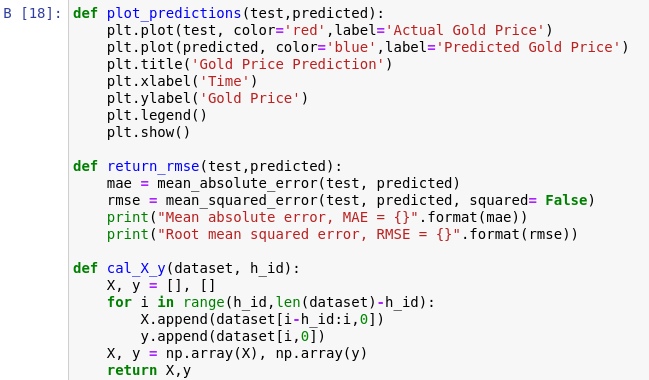


Рисунок 3.2.3 Фрагмент кода для графика на рисунке 3.1.3, средней абсолютной ошибки и для средней квадратичной ошибки.

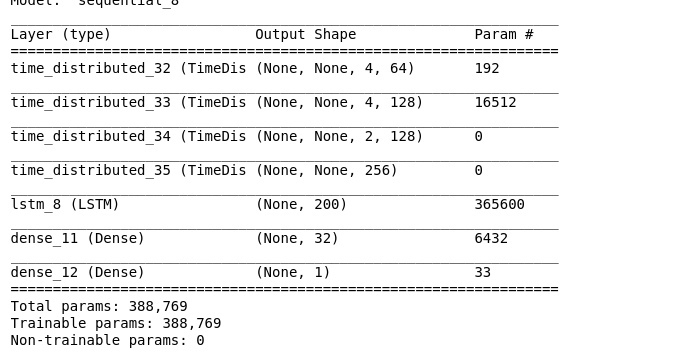


Рисунок 3.2.4 Модель номер два.

### 3.3. Тестирование программы

В качестве проверки работоспособности системы результаты будут показаны на графике где будет показана стоимость реальная стоимость на рынке и стоимость которую предсказала нейронная сеть.

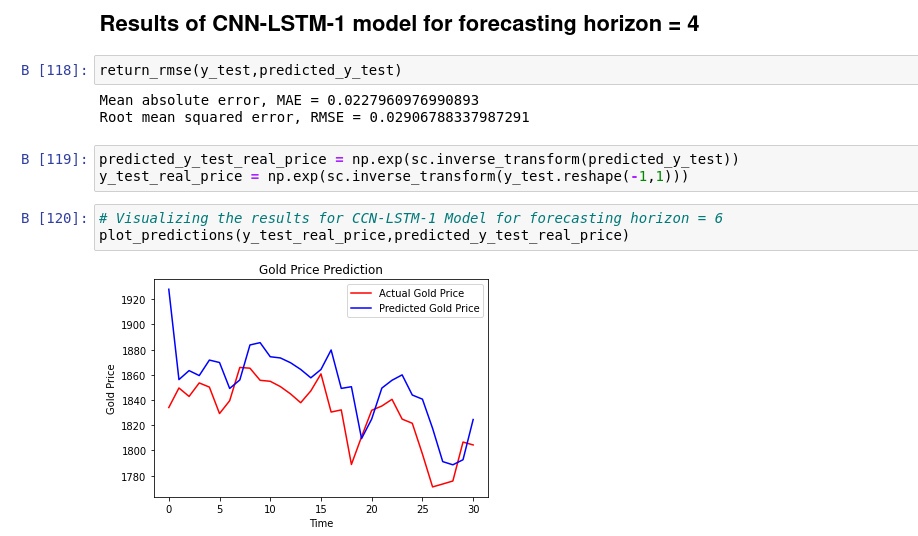


Рисунок 3.3.1 Результаты для первой модели.

При горизонте прогнозирования в 4 дня первая модель своим предсказанием выдает сильно завышенный показатель стоимости, вероятно это происходит из за того, что в предыдущие дни стоимость достаточно стремительно росла и нейронная сеть посчитала, что рост продолжиться. Но в дальнейшем графики начинают сходиться и хоть и имеется погрешность в целом 10 долларов за унцию, но динамику падения и роста цен нейронная сеть предсказала правильно.

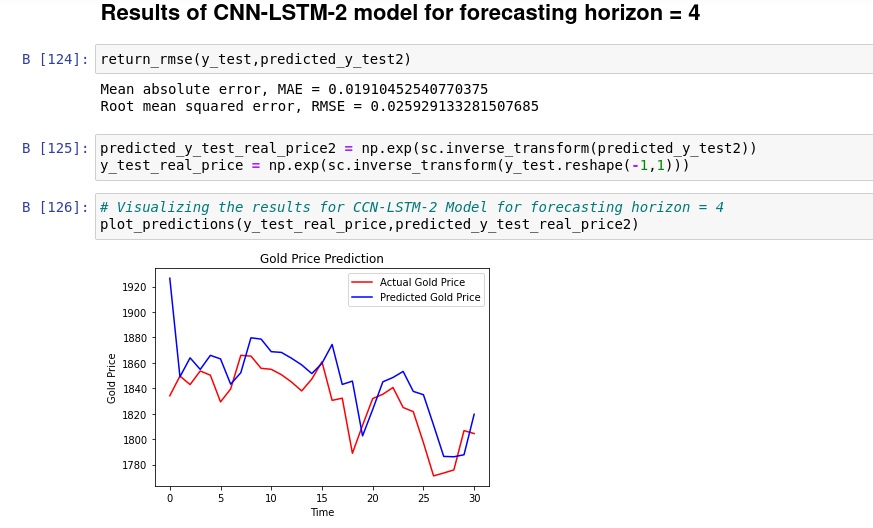


Рисунок 3.3.2 Результаты модели номер два.

На второй модели результаты сильно похожи и можно сказать что результаты получились лучше, так как графики сошлись еще больше и погрешность становиться несколько меньше. Динамика роста и падения цен также исправно соблюдается.

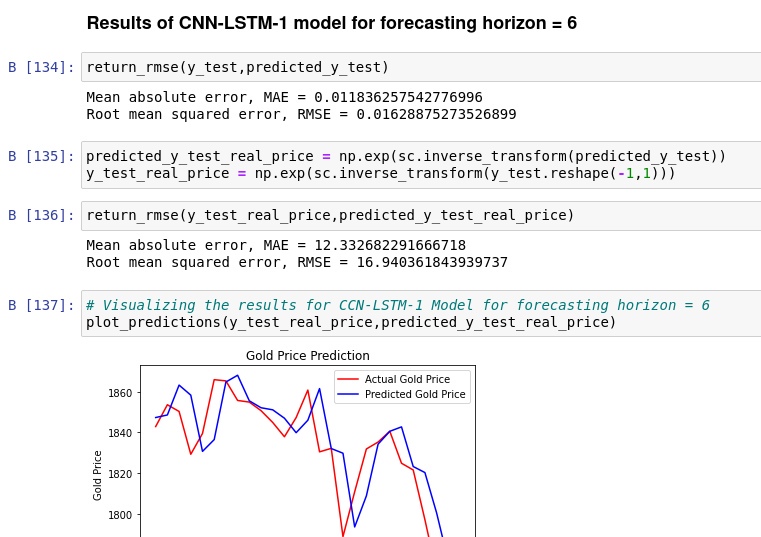


Рисунок 3.3.3 Результаты модели номер один.

На данном графике получились одни из самых лучших результатов, с начальной стоимостью погрешность все также значительно, но за тем нейронная сеть очень точно попадает в локальные максимумы и минимумы стоимости, также стоит заметить то, что графики практически идентичны и с помощью данной модели вполне возможно отслеживать динамику.

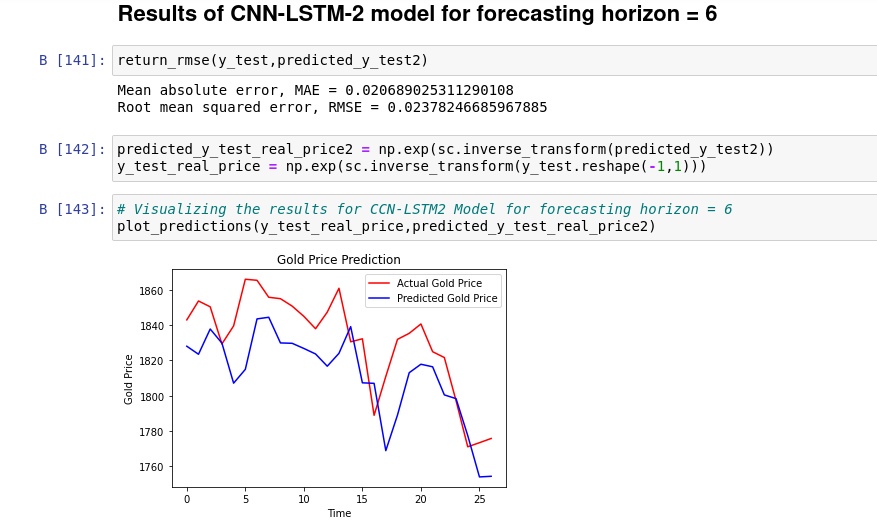


Рисунок 3.3.4 Результаты модели номер два.

Данная модель, на данном датасете, как по мне, абсолютно неудачна, так как значения вначале временной шкалы имеют слишком большой разброс как по стоимости так и графики не сошлись, динамику в данном случае предсказать будет проблематично.

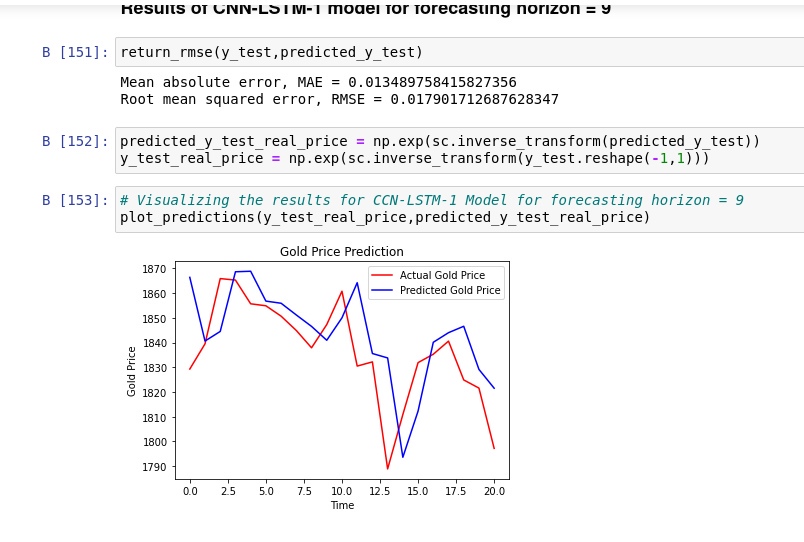


Рисунок 3.3.5 Результаты модели номер один.

На данном рисунке проблема первого предсказания также очень заметна, но затем получаются весьма точными, эта модель имеет право на существование и дальнейшее ее использование на долгосрочной перспективе

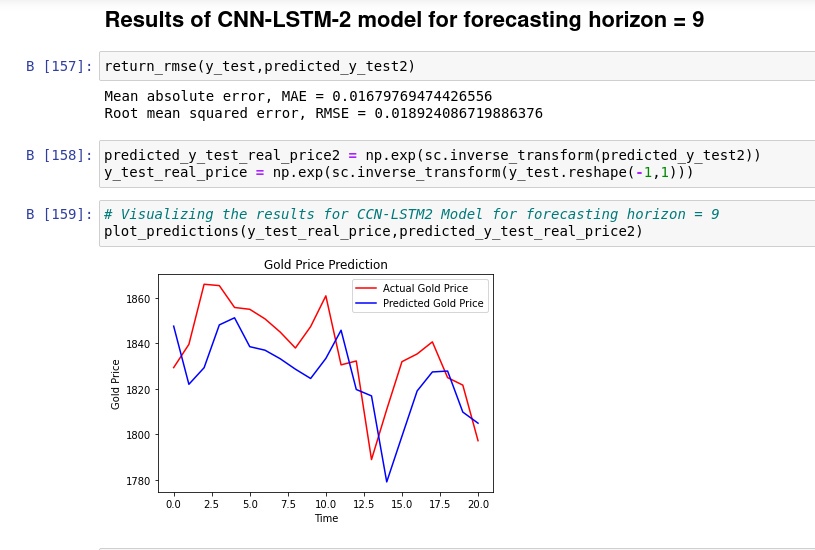
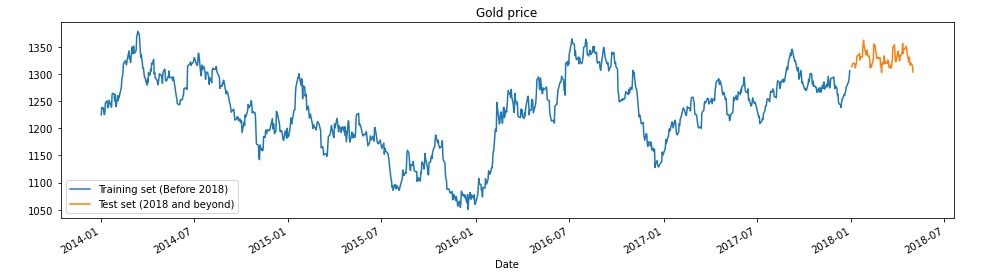


Рисунок 3.3.6 Результаты модели номер два.

Первая модель на данном горизонте прогнозирования отработала немного лучше, но результаты в целом схожи.

### 3.4 Тестирование программы на других временных периодах



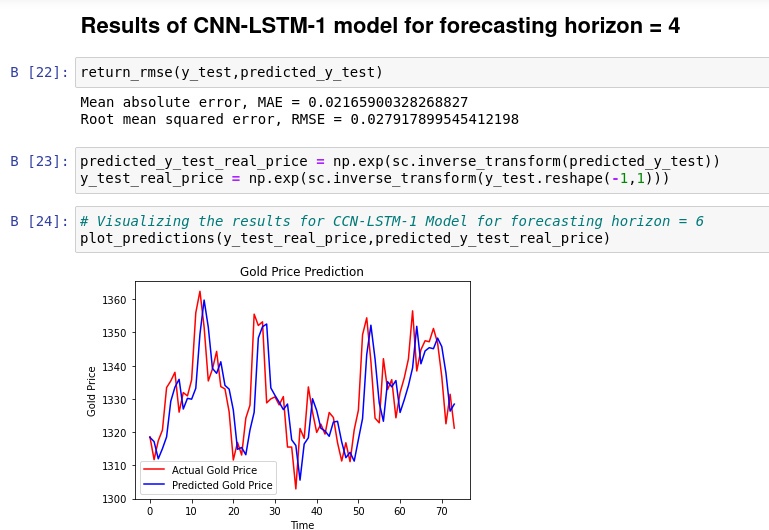


Рисунок 3.4.1 На данном графике ясно видно, что на данном временном промежутке нейросеть ведет себя намного лучше, вероятно это связано с большей стабильностью в те года, то есть “неосязаемые” факторы вносят не такой большой вклад в изменения.

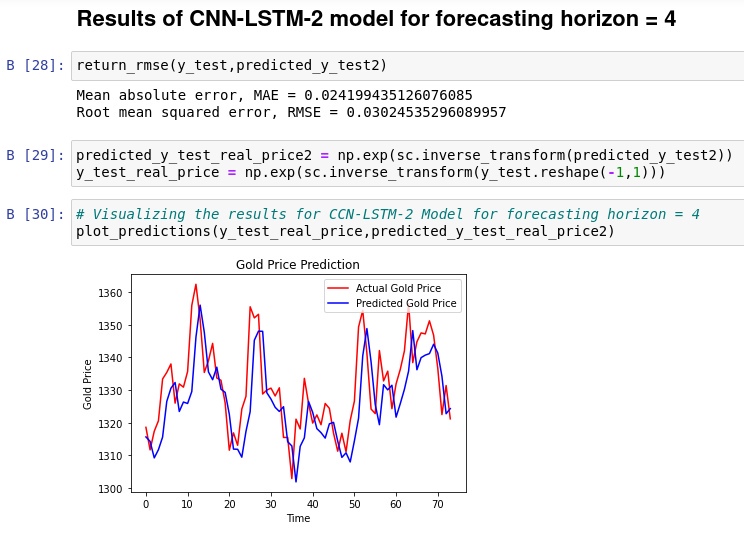
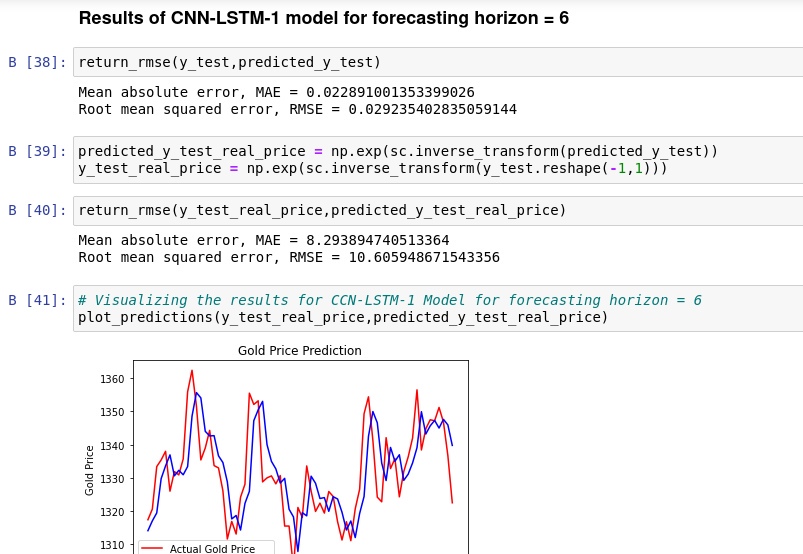


Рисунок 3.4.2 Здесь ситуация похожая, но локальные минимумы посчитаны хуже.



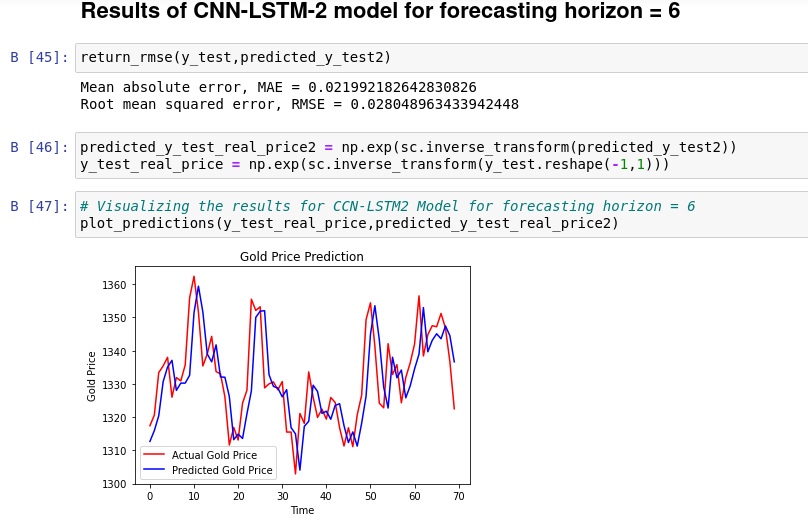
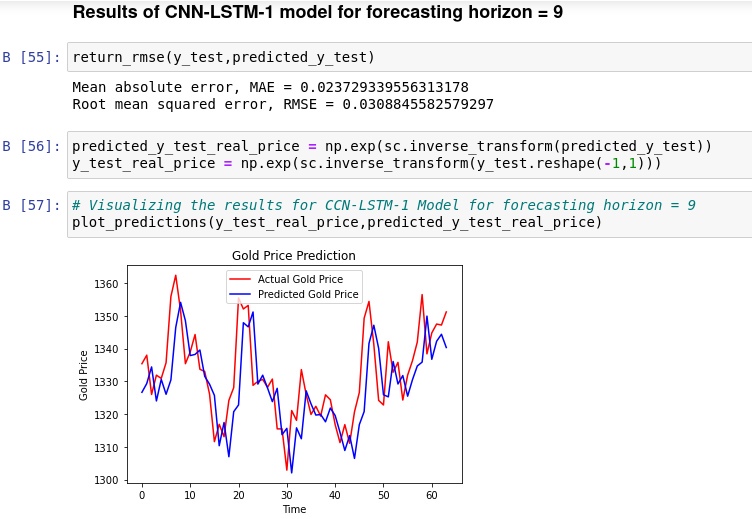


Рисунок 3.4.3 На данном рисунке видно что нейронная сеть продолжает исправно выдавать правильные предсказания даже при горизонте прогнозирования в 9 дней. Но стоит заметить что пиковые значения все же расходятся достаточно значительно.



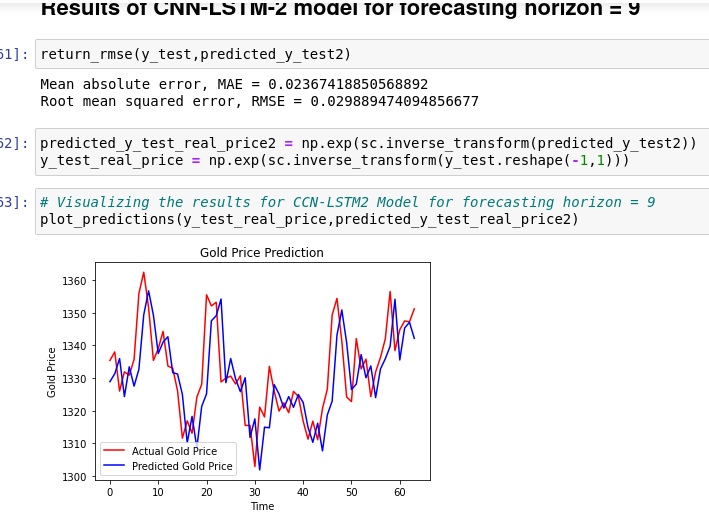


Рисунок 3.4.4 На данном рисунке видно что нейронная сеть продолжает исправно выдавать правильные предсказания даже при горизонте прогнозирования в 6 дней.

### 

### Заключение

Согласно хронологическим характеристикам данных о ценах на акции, в этой работе предлагается CNN-LSTM для прогнозирования цены закрытия акций. В качестве входных данных метод использует цену открытия, максимальную цену, минимальную цену, цену закрытия, объем, оборот, взлеты и падения, а также изменение данных по акциям, в полной мере использовать характеристики временной последовательности данных о запасах. CNN используется для извлечения характеристик входных данных. LSTM используется для изучения извлеченных данных характеристик и прогнозирования цены закрытия акции. В этой работе в качестве примера для проверки экспериментальных результатов используются соответствующие данные GF-21.

Результаты экспериментов показывают, что CNN-LSTM имеет самую высокую точность прогнозирования и лучшую производительность по сравнению с MLP, CNN, RNN, LSTM и CNN-RNN. CNN-LSTM подходит для прогнозирования цен на акции и может служить релевантной справочной информацией для инвесторов, чтобы максимизировать отдачу от инвестиций. CNN-LSTM также предлагает практический опыт исследования данных финансовых временных рядов. Однако у модели все же есть недостатки. Например,

он учитывает только влияние данных о ценах акций на цены закрытия и не учитывает эмоциональные факторы, такие как новости и национальная политика, в прогнозе.

### 

### Список использованных источников

1. Люгер, Дж.Ф. Искусственный интеллект: стратегии и методы решения

сложных проблем / Дж.Ф. Люгер – М.: Вильямс, 2003. – 864 с.

1. S.Haykin. Neural Networks and Learning Machines. 3rd Edition. Pearson, 2018. – 944 с.
2. Ф.М. Гафаров, А.Ф.Галимянов. Искусственные нейронные сети и их приложения, – Казань:КФУ, 2018. – 121 с.
3. Тарик Рашид. Make Your Own Neural Network, 2016. – М.: Вильямс, 2003. – 222 с.