

Правительство Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»
(НИУ ВШЭ)

Московский институт электроники и математики им. А.Н. Тихонова

ПРАКТИЧЕСКАЯ РАБОТА № 12
по дисциплине «Информатика»

ТЕМА РАБОТЫ

«Анализ и прогнозирование временных рядов с применением ARIMA и
Holt-Winters»

Москва, 2024

Оглавление

1. Введение	2
2. Содержание практической работы	4
3. Ход работы	5
4. Приобретаемые навыки	20
5. Обобщенная задача для индивидуального варианта	20
6. Распределение вариантов	22

1. Введение

Целью данной лабораторной работы является освоение методов анализа временных рядов и прогнозирования данных с использованием моделей ARIMA и Holt-Winters в инструменте RapidMiner. Эти модели широко применяются для анализа данных, демонстрирующих тренды и сезонные колебания, что делает их особенно полезными в задачах прогнозирования.

В рамках работы студенты выполняют загрузку и предобработку данных о временных рядах цен акций MSFT, обучают модели ARIMA и Holt-Winters, а также проведут сравнительный анализ их прогнозов. Особое внимание будет уделено визуализации результатов, которая позволит наглядно оценить точность предсказаний обеих моделей.

Кроме того, студенты изучат методы оценки качества прогнозов с помощью визуального анализа графиков реальных и предсказанных значений, а также научатся интерпретировать полученные результаты.

2. Содержание практической работы

Описание работы:

В основе работы лежит анализ временных рядов стоимости акций Microsoft (MSFT) с применением методов прогнозирования ARIMA и Holt-Winters. Данные включают ежедневные значения цен закрытия с 09.03.2020 по 07.03.2025. Основная цель работы – построение и сравнение моделей прогнозирования стоимости акций на основе исторических данных.

Этапы выполнения работы:

1. Загрузить данные о временных рядах стоимости акций MSFT
2. Провести предобработку данных, задав необходимые роли атрибутам
3. Построить модель прогнозирования ARIMA
4. Построить модель прогнозирования Holt-Winters
5. Применить обученные модели для прогнозирования будущих значений стоимости акций
6. Визуализировать результаты прогнозирования обеих моделей
7. Сравнить точность предсказаний моделей ARIMA и Holt-Winters

О наборе данных:

Анализ проводится на наборе данных, содержащем следующие характеристики:

- **Date** — дата торгов
- **Price close** — значение стоимости акции на момент закрытия торгов

Ключевые особенности данных:

Количество записей: содержит данные о финальной стоимости акций Microsoft для 1257 дней торгов.

Формат данных: временной ряд с дневной периодичностью, числовые данные.

Потенциальные проблемы:

Наличие трендов и сезонных колебаний в данных, также возможны выбросы из-за рыночных аномалий. Требуется правильный подбор параметров моделей.

Специфика работы:

Для выполнения лабораторной работы используются все данные представленные в датасете.

- **Date** (как временная метка),
- **Price close** (как целевой признак для прогнозирования)

Этот набор данных позволяет провести детальный анализ временных рядов и построить модели прогнозирования для краткосрочного и среднесрочного предсказания цен на акции Microsoft.

3. Ход работы

Загрузка набора данных

1. Откройте RapidMiner Studio.
2. В главном меню выберите "**Create New Process**".
3. Воспользуйтесь функцией "**Import data**".
4. Загрузите набор данных о транзакциях, выбрав файл "**MSFT_data**" в формате csv.
5. Убедитесь, что для столбцов верно указаны типы данных – Date для временных атрибутов и real для финальной стоимости соответственно.
6. Сохраните полученную базу данных в папку со своей работой.
7. В результате вы увидите таблицу с данными, содержащими атрибуты: (**Date, Price close**)

Данные успешно загружены, их структура показана на рисунке 3.2.

Import Data - Specify your data format

Specify your data format

☒ Header Row

1

Start Row

1

Column Separator

Semicolon ";"

File Encoding

windows-1251

Escape Character

\

Decimal Character

.

☒ Use Quotes

"

☒ Skip Comments

#

☐ Trim Lines

☐ Multiline Text

1	Date	Price Close
2	09.03.2020	144.17025756835938
3	10.03.2020	154.02919006347656
4	11.03.2020	147.05137634277344
5	12.03.2020	133.10528564453125
6	13.03.2020	152.0286865234375
7	16.03.2020	129.6211395263672
8	17.03.2020	140.29368591308594
9	18.03.2020	134.3878631591797
10	19.03.2020	136.5989990234375
11	20.03.2020	131.468505859375
12	23.03.2020	130.15716552734375
13	24.03.2020	141.9878692626953
14	25.03.2020	140.62869262695312
15	26.03.2020	149.42520141601562
16	27.03.2020	143.2896728515625
17	30.03.2020	153.36874389648438
18	31.03.2020	150.956656730625

no problems.

Previous Next Cancel

Рисунок 3.1 – подготовка данных к выгрузке

Import Data - Format your columns.

Format your columns.

Date format

dd.MM.yyyy

100%

☐ Replace errors with missing values

	Date	Price Close
	date	real
1	Mar 9, 2020	144.170
2	Mar 10, 2020	154.029
3	Mar 11, 2020	147.051
4	Mar 12, 2020	133.105
5	Mar 13, 2020	152.029
6	Mar 16, 2020	129.621
7	Mar 17, 2020	140.294
8	Mar 18, 2020	134.388
9	Mar 19, 2020	136.599
10	Mar 20, 2020	131.469
11	Mar 23, 2020	130.157
12	Mar 24, 2020	141.988
13	Mar 25, 2020	140.629
14	Mar 26, 2020	149.425
15	Mar 27, 2020	143.290
16	Mar 30, 2020	153.369
17	Mar 31, 2020	150.957
18	Apr 1, 2020	145.596

no problems.

Previous Next Cancel

Рисунок 3.2 – выгруженные данные

Назначение ролей:

После того как данные загружены в RapidMiner, необходимо правильно назначить роли атрибутам. Для этого в проект добавляется оператор **"Set Role"**. Этот блок используется для изменения ролей атрибутов, чтобы указать, какие из них будут использоваться как целевые переменные (что прогнозируем) и какие будут служить временными метками или другими характеристиками. Атрибуту Date присваивается metadata, а Price close получает роль Label и таким образом становится целевой переменной как показано на рисунке 3.3

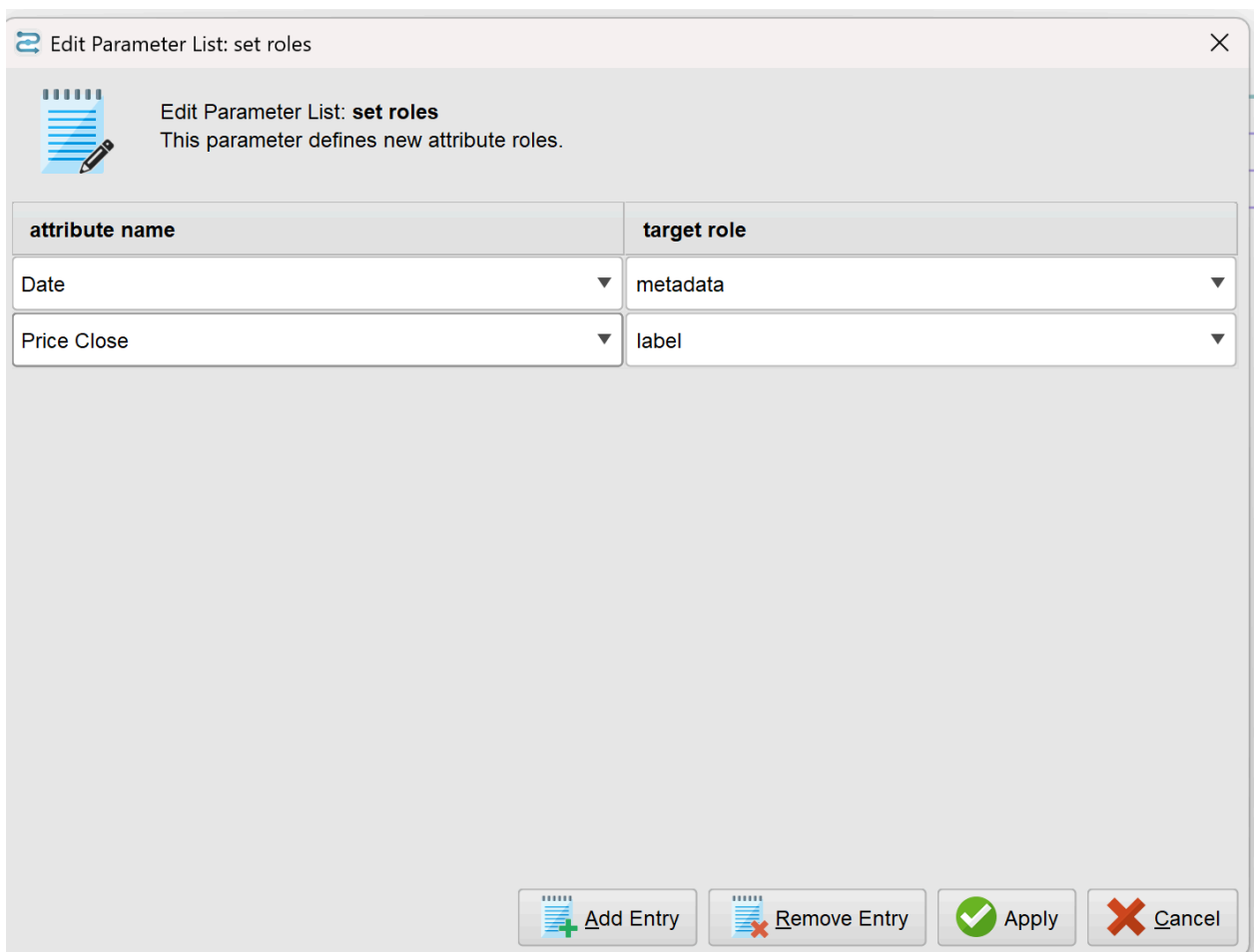


Рисунок 3.3 – настройки для Set Roles

Блок Arima:

Блок используется для построения модели прогнозирования временных рядов с помощью метода ARIMA (AutoRegressive Integrated Moving Average).

Параметры блока:

- **Time Series Attribute (Price Close):** Указывает на атрибут Price Close для прогнозирования.

- **Has Indices:** Указывает, что данные имеют индексы (временные метки).
- **p:** Порядок авторегрессионной модели (AR), установлено **p = 1**, что означает использование одного предыдущего значения для прогноза.
- **d:** Степень дифференцирования (0), указывает, что данные уже стационарны.
- **q:** Порядок модели скользящего среднего (MA), установлено **q = 1**, что означает использование одного лагового значения ошибки.
- **Estimate Constant:** Включает оценку постоянной величины в модели.
- **Main Criterion (AIC):** Для выбора наилучшей модели используется AIC (Akaike Information Criterion).
- **Error Handling:** Выбрана опция **use default forecast** для стандартного прогнозирования.

Финальные настройки блока представлены на рисунке 3.4.

Parameters

ARIMA

time series attribute
Price Close

☐ has indices

p: order of the autoregressive model
1

d: degree of differencing
0

q: order of the moving-average model
1

☒ estimate constant

main criterion
aic

error handling
use default forecast...

[Show advanced parameters](#)

Рисунок 3.4 – настройки для Arima

Блок Holt-Winters:

Блок используется для прогнозирования временных рядов с учетом сезонности и тренда. Этот метод позволяет моделировать данные, где наблюдается как долгосрочный тренд, так и регулярные сезонные колебания.


Параметры блока:

- **Time Series Attribute (Price Close):** Атрибут **Price Close** выбран как целевая переменная для прогнозирования.

- **Alpha:** Коэффициент сглаживания уровня (0.5). Этот параметр управляет тем, насколько сильно сглаживаются данные для уровня.
- **Beta:** Коэффициент сглаживания тренда (0.1). Этот параметр регулирует, насколько сильно модель реагирует на изменения тренда в данных.
- **Gamma:** Коэффициент сглаживания сезонности (0.5). Указывает, насколько модель будет учитывать сезонные колебания данных. Значение 0.5 указывает на умеренное влияние сезонности.
- **Period:** Длительность одного периода (4). Этот параметр определяет количество наблюдений, которые составляют один период. В данном случае это означает, что сезонность моделируется с периодом в 4 единицы времени (в рамках данной работы - 4 дня).
- **Seasonality Model:** Модель сезонности (multiplicative). Это означает, что модель предполагает, что сезонные колебания умножаются на тренд, а не добавляются к нему.

Финальные настройки блока показаны на рисунке 3.5.

Parameters

 Holt-Winters

time series attribute

Price Close

☐ has indices

alpha: coefficient for level smoothing

0.5

beta: coefficient for trend smoothing

0.1

gamma: coefficient for seasonality smoothing

0.5

period: length of one period

4

seasonality model

multiplicative


 [Show advanced parameters](#)

Рисунок 3.5 – настройки для Holt-Winters

Применение прогнозов:

Для применения построенной модели прогнозирования используется блок "Apply Forecast". Этот блок позволяет сгенерировать предсказания для временного ряда, используя обученную модель.

Параметры блока Apply Forecast:

- **Forecast Horizon:** Параметр **forecast horizon** определяет, на сколько шагов вперед модель будет делать прогноз. В данном случае

установлено **5**, что означает, что прогноз будет рассчитан на 5 будущих периодов (5 дней в рамках данной работы).

- **Add Original Time Series:** Этот параметр позволяет добавить оригинальный временной ряд в выходные данные. Это полезно, если вы хотите сравнить реальные значения с прогнозируемыми, так как все данные (реальные и прогнозируемые) будут отображены в одном выходном потоке.
- **Add Combined Time Series:** Этот параметр позволяет добавить комбинированный временной ряд, включающий как фактические значения, так и прогнозируемые. Это поможет увидеть, как прогнозы соотносятся с реальными данными на одном графике или в одном наборе данных.

Настройки для блоков Apply Forecast для моделей идентичны и представлены на рисунке 3.6.

Parameters ✕

Apply Forecast (2) (Apply Forecast)

forecast horizon ⓘ

☒ add original time series ⓘ

☒ add combined time series ⓘ

[Show advanced parameters](#)

Рисунок 3.6 – параметры для Decision tree

Также перед тем, как запустить проект, был добавлен блок **"Statistics"**, который позволяет провести предварительный анализ данных. Этот оператор автоматически вычисляет ключевые статистические показатели для каждого атрибута в наборе данных, включая минимальное, максимальное и среднее значение, а также стандартное отклонение. Финальный вид схемы и корректное подключение блоков представлены на рисунке 3.7.

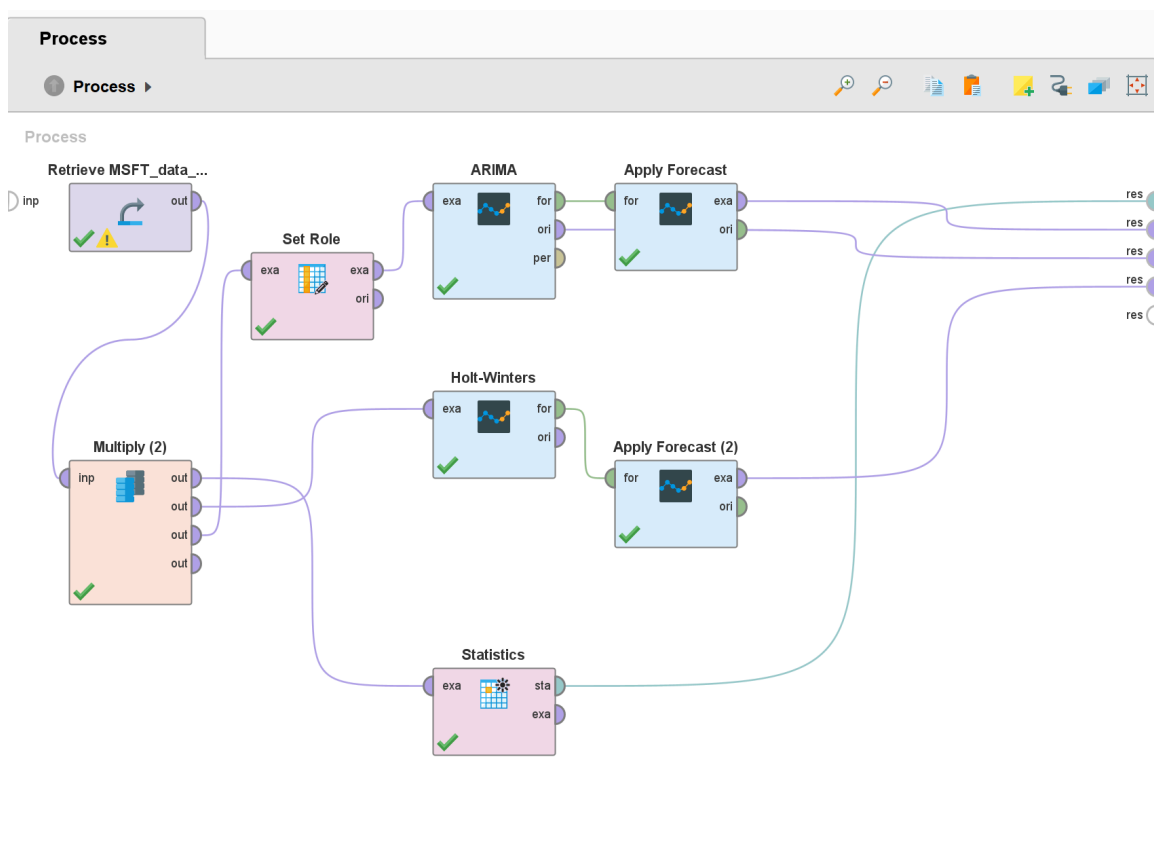


Рисунок 3.7 – Финальный вид схемы

После запуска процесса мы получили несколько результатов, представленных на рисунке 3.8. В частности, был создан "**ExampleSet (Set Role)**" (представленный на рисунке 3.9), который корректно определил параметры атрибутов для времени и стоимости, об этом свидетельствует выделение цветом соответствующих столбцов.

С помощью функции **Statistics** в этом же блоке, можем просмотреть основные результаты анализа, который провёл блок. Среди них – пропуски, минимальное и максимальное значения, среднее значение, а также отклонение, все результаты показаны на рисунке 3.10.

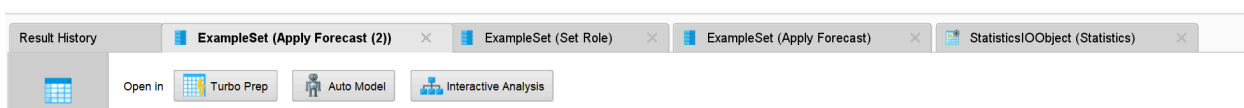


Рисунок 3.8 – Полученные результаты

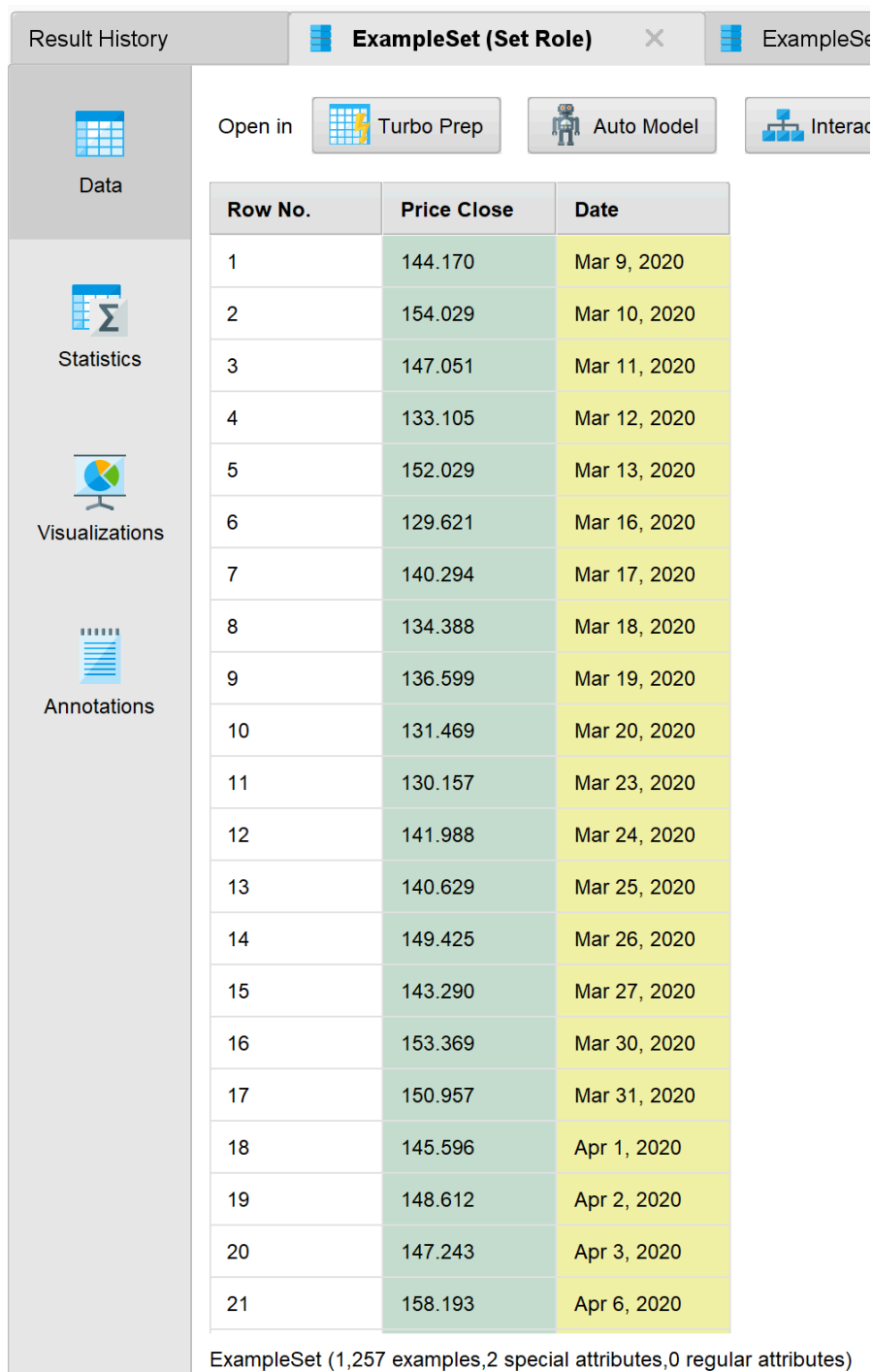


Рисунок 3.9 – Example set

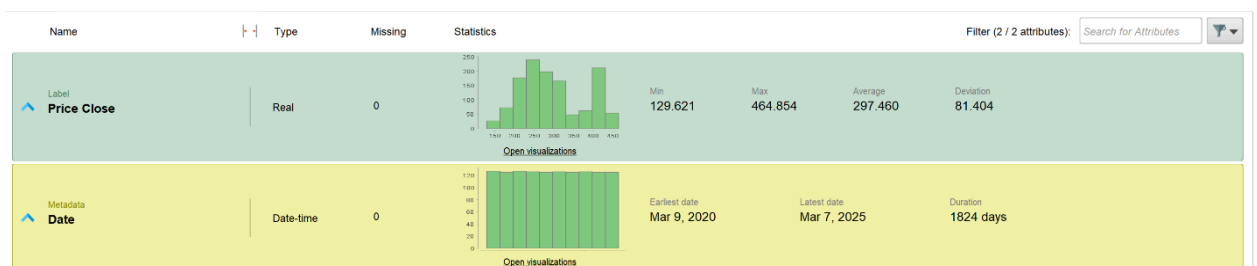


Рисунок 3.10 – статистика для датасета

Результаты прогнозирования модели ARIMA:

Теперь можно перейти к анализу результатов, полученных после использования модели. Основными результатами прогнозирования стали 5 значений, которые представлены на рисунке 3.11. Они отображены в строках с 1258 по 1262, поскольку уже выходят за временные рамки, указанные в датасете.

Для удобства анализа можно воспользоваться вкладкой "Visualizations", которая находится в левом меню интерфейса. Здесь можно выбрать тип графика (Plot type), указать оси и сгруппировать данные. В данном случае был выбран график (Spline), где по оси X указаны дни измерений, а по оси Y – финальная стоимость акций. Результаты представлены на рисунке 3.12.

1257	?	393.310	393.310
1258	393.611	?	393.611
1259	393.581	?	393.581
1260	393.550	?	393.550
1261	393.520	?	393.520
1262	393.490	?	393.490

ExampleSet (1,262 examples,1 special attribute,2 regular attributes)

Рисунок 3.11 – результаты прогнозирования

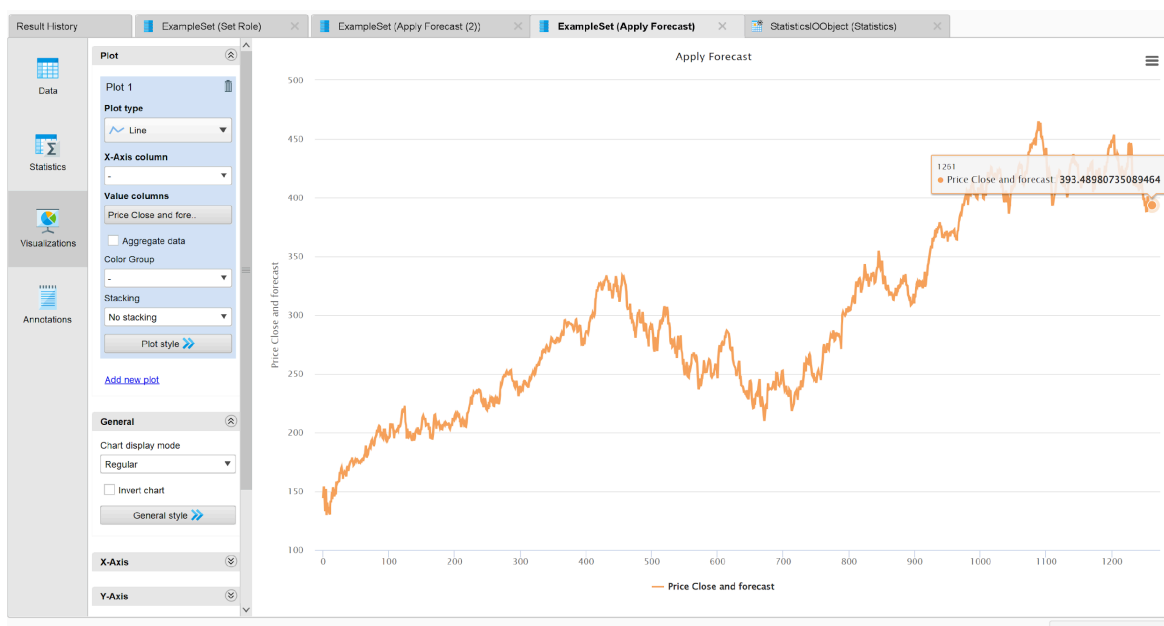


Рисунок 3.12 – визуализация прогнозирования

Проведём аналогичные действия для данных, полученных с помощью прогнозирования с помощью модели Holt-winters. Таблица с результатами представлен на рисунке 3.13. В строках с прогнозами, 1258-1262, можно увидеть значения, которые система предсказала на основе модели Holt-Winters. Прогнозы отображают значения, которые постепенно увеличиваются, отражая тренд роста в данных.

Также аналогичным образом с помощью функции "Visualizations" построим график. Результаты показаны на рисунке 3.14. Holt-Winters дает более сглаженные прогнозы, чем ARIMA, из-за учета сезонности и тренда. Однако, в момент более резких изменений на графике, модель продолжает следовать за основной тенденцией, что подтверждает её способность захватывать долгосрочные тренды.

1256	?	396.890	396.890
1257	?	393.310	393.310
1258	394.835	?	394.835
1259	396.196	?	396.196
1260	393.689	?	393.689
1261	389.827	?	389.827
1262	391.717	?	391.717

ExampleSet (1,262 examples,1 special attribute,2 regular attributes)

Рисунок 3.13 – результаты прогнозирования

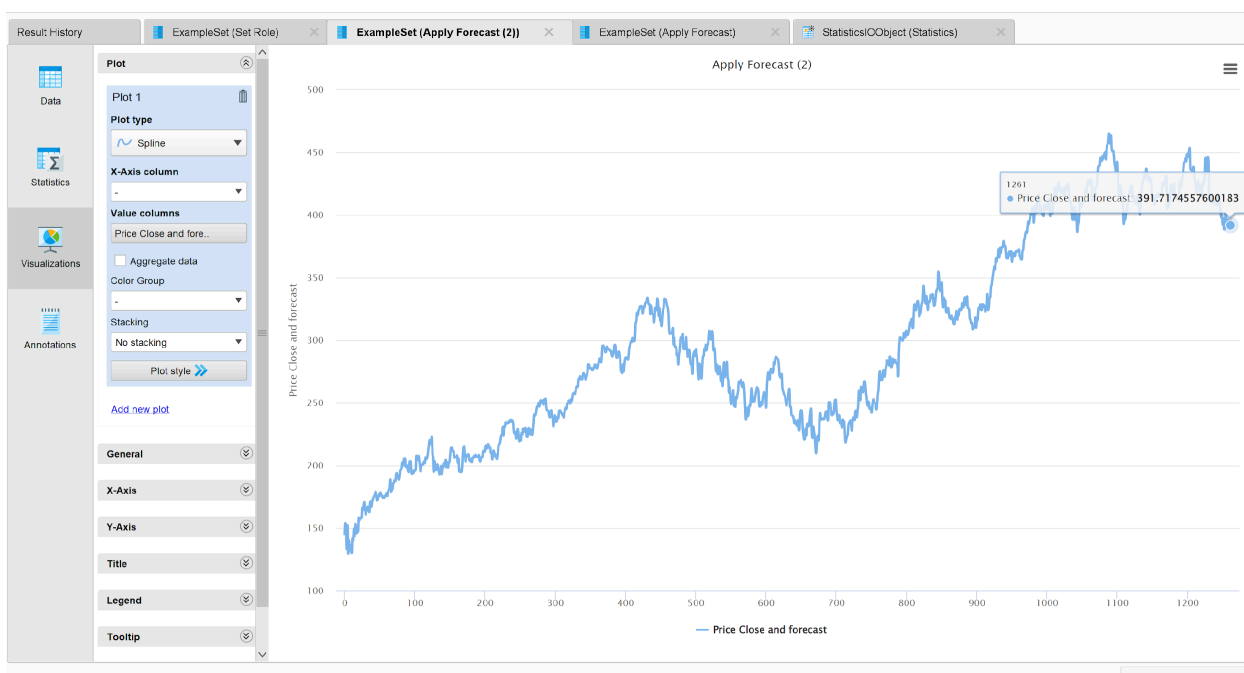


Рисунок 3.14 – визуализация полученных результатов

Общие выводы:

Обе модели — ARIMA и Holt-Winters — достаточно хорошо справились с прогнозированием цен на акции MSFT, учитывая тренды и сезонность. Прогнозы, полученные для периода с 10 по 14 марта 2025 года, соответствуют ожидаемым значениям и показывают разумную динамику,

хотя и с некоторыми отклонениями от реальных данных. Реальная стоимость акций подставлена в таблице 3.1

Дата	Стоимость
10.03.2025	380.1600036621094
11.03.2025	380.45001220703125
12.03.2025	383.2699890136719
13.03.2025	378.7699890136719
14.03.2025	388.55999755859375

Таблица 3.1 – реальная стоимость акций в прогнозируемый период

Тем не менее, стоит отметить, что прогнозирование рынка акций — это сложная задача, поскольку стоимость акций зависит от множества факторов, новостных, экономических событий, изменений в бизнесе компаний и других внешних воздействий. Даже самые передовые модели могут не всегда точно предсказать движение цен, особенно в условиях высокой волатильности и нестабильности рынка.

Таким образом, модели показали хорошие результаты, но для более точных прогнозов важно учитывать дополнительные данные и факторы, которые могут существенно влиять на поведение рынка.

4. Приобретаемые навыки

1. Работа с интерфейсом RapidMiner Studio – освоение инструментов для построения процессов, анализа данных и настройки операторов.
2. Построение моделей прогнозирования – использование моделей ARIMA и Holt-Winters для предсказания. Пользователь освоит настройку параметров моделей для получения более точных прогнозов проанализировав их работу на реальных данных.
3. Применение модели к тестовым данным – освоение оператора Apply Forecast для применения моделей на тестовых данных и корректная интерпретация полученных прогнозов.
4. Развитие навыков анализа данных – интерпретация полученных результатов, выявление закономерностей и подготовка отчёта по итогам лабораторной работы.

5. Обобщенная задача для индивидуального варианта

Цель работы – прогнозирование временных рядов с помощью классических моделей ARIMA и модели Хольта-Уинтерса. Выполните в RapidMiner (Altair AI Studio) следующие шаги:

- 1) Загрузка и подготовка данных
 - Импортируйте CSV-файл с временным рядом и назначьте роль атрибутам: временная метка → date (табличный индекс или date-тип), измерение → label.
 - Проверьте и при необходимости устраните пропуски (Replace Missing Values), откройте сезонные разладки (Generate Attributes → Dummy для праздников/выходных), приведите периодичность данных к нужной

(Fill Missing Values → linear/interpolate).

2) Построение и настройка моделей

ARIMA:

- Подберите порядок (p,d,q) с помощью автоматической оценки (AIC/BIC) или вручную;
- Оцените параметры модели и выполните прогноз на заданный горизонт.

Holt–Winters:

- Выберите аддитивную или мультипликативную модель в зависимости от характера сезонности;
- Задайте коэффициенты сглаживания уровня, тренда и сезонности (α , β , γ);
- Постройте прогноз.

3) Прогнозирование и визуализация

- С помощью оператора Apply Forecast сгенерируйте прогноз для заданного числа периодов вперёд.
- Постройте на одном графике исторические данные и предсказания обеих моделей (Line/Spline Chart).

4) Оценка качества и сравнение

- Рассчитайте метрики ошибки прогноза (MAE, RMSE, MAPE) через оператор Performance (Forecast).
- Сравните модели по метрикам и визуальному соответствию реальным значениям.

5) Выводы

- Определите, какая модель лучше справляется с учетом тренда и сезонности в вашем конкретном ряде.

6. Распределение вариантов

