

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ»

КАФЕДРА БИЗНЕС-ИНФОРМАТИКИ И МЕНЕДЖМЕНТА

ОТЧЕТ
ЗАЩИЩЕН С ОЦЕНКОЙ
РУКОВОДИТЕЛЬ

канд. техн. наук, доцент
должность, уч. степень, звание

подпись, дата

А.Д. Державина
инициалы, фамилия

ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ
ПОДБОР ДАННЫХ ДЛЯ РЕШЕНИЯ ЗАДАЧ
ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА
И ЭЛЕМЕНТЫ ПРОГРАММИРОВАНИЯ В RAPID MINER

по дисциплине: МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ ПРИНЯТИЯ РЕШЕНИЙ

РАБОТУ
ВЫПОЛНИЛ
СТУДЕНТ ГР. №

8426МК

подпись, дата

Э.В. Пекарская
инициалы, фамилия

Санкт-Петербург 2024

Цель работы: получение начальных навыков работы с программным комплексом RapidMinerStudio, сформировать модель регрессии, подготовленной разработчиками пакета RM, и для собственного набора данных.

Ход работы

В RapidMiner был разобран пример применения метода линейной регрессии к заданному набору данных LinearRegression. Созданная модель линейной регрессии, позволяет делать численные прогнозы для рассматриваемого набора данных. Процесс получения результатов представлен на рисунках 1 – 3.

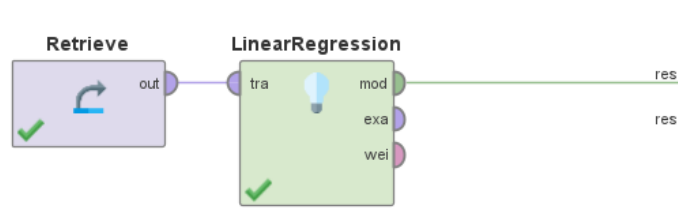


Рисунок 1 – Модель линейной регрессии

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
a1	31.736	2.114	0.463	1.000	15.009	0	****
a2	42.948	1.928	0.689	0.998	22.276	0	****
a3	23.773	2.000	0.369	1.000	11.888	0	****
a4	3.706	2.060	0.056	1.000	1.799	0.074	*
a5	-4.184	1.923	-0.068	0.999	-2.176	0.031	**
(Intercept)	-304.228	23.603	?	?	-12.889	0	****

Рисунок 2 – Результаты моделирования

LinearRegression

```

31.736 * a1
+ 42.948 * a2
+ 23.773 * a3
+ 3.706 * a4
- 4.184 * a5
- 304.228

```

Рисунок 3 – Описание линейной регрессии

LinearRegression (линейная регрессия) — это алгоритм машинного обучения, который используется для прогнозирования численных значений на

основе линейной зависимости между независимыми переменными (факторами) и зависимой переменной (целевым значением).

В RapidMiner, LinearRegression является одним из множества операторов, эффективно реализующих алгоритм линейной регрессии. Он используется для построения модели линейной регрессии на тренировочных данных, а затем применения этой модели для прогнозирования значений на тестовых данных.

Далее был выбран и загружен собственный набор данных – котировки акций МОЕХ Сбербанк и IMOEX Московская Биржа за период с 15 сентября 2022 года по 15 сентября 2023 года. С помощью метода полиномиальной регрессии (Polynomial Regression) удалось построить и исследовать модель полиномиальной регрессии.

Чтобы иметь возможность визуализации процесса и оценки разработанной модели, данные были разделены на тестовый и обучающий наборы с помощью оператора Split Data и использован оператор Apply Model. В режиме Visualizations построен график, который демонстрирует совпадение исходных и предсказанных данных модели.

Результаты представлена на рисунках 4 – 7.

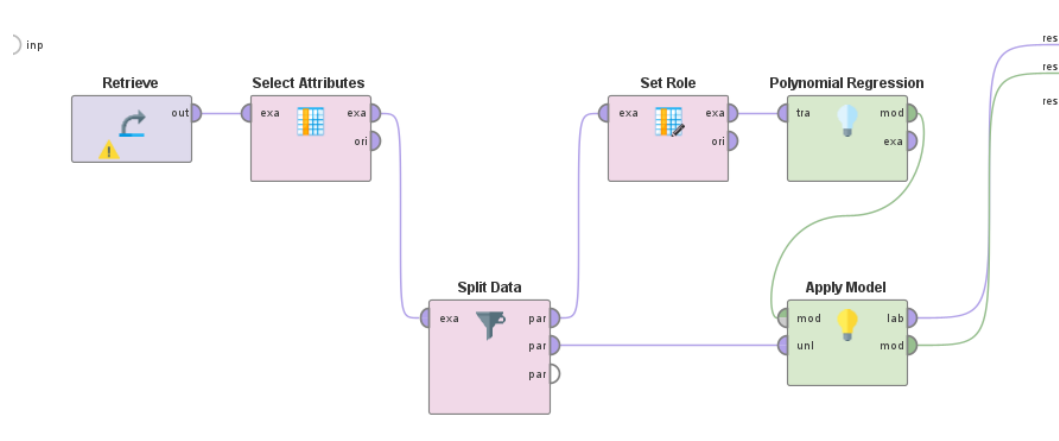


Рисунок 4 – Модель процесса вычисления полиномиальной регрессии

Row No.	<DATE>	IMOEX Мос. Биржа	prediction(MOEX СберБанк)	MOEX СберБанк
25	Dec 29, 2022	2147.060	161.684	139.210
26	Dec 30, 2022	2154.120	162.327	140.260
27	Jan 4, 2023	2168.420	163.631	140.160
28	Jan 9, 2023	2163.500	163.183	141.100
29	Jan 18, 2023	2196.260	166.169	152.250
30	Jan 20, 2023	2166.690	163.473	151.600
31	Jan 24, 2023	2172.750	164.026	153.740
32	Jan 25, 2023	2170.150	163.789	152.770
33	Jan 30, 2023	2204.410	166.912	153.450
34	Jan 31, 2023	2225.600	168.844	155.220
35	Feb 2, 2023	2243.540	170.480	161.240
36	Feb 6, 2023	2272.370	173.108	166.890
37	Feb 7, 2023	2267.690	172.681	164.900
38	Feb 9, 2023	2262.450	172.204	165.970
39	Feb 13, 2023	2264.270	172.370	164.680
40	Feb 15, 2023	2166.460	163.452	156.050

Рисунок 5 – Результат моделирования

PolynomialRegression

0.091 * IMOEX Мос. Биржа ^ 1.000
- 34.061

Рисунок 6 – Описание регрессии

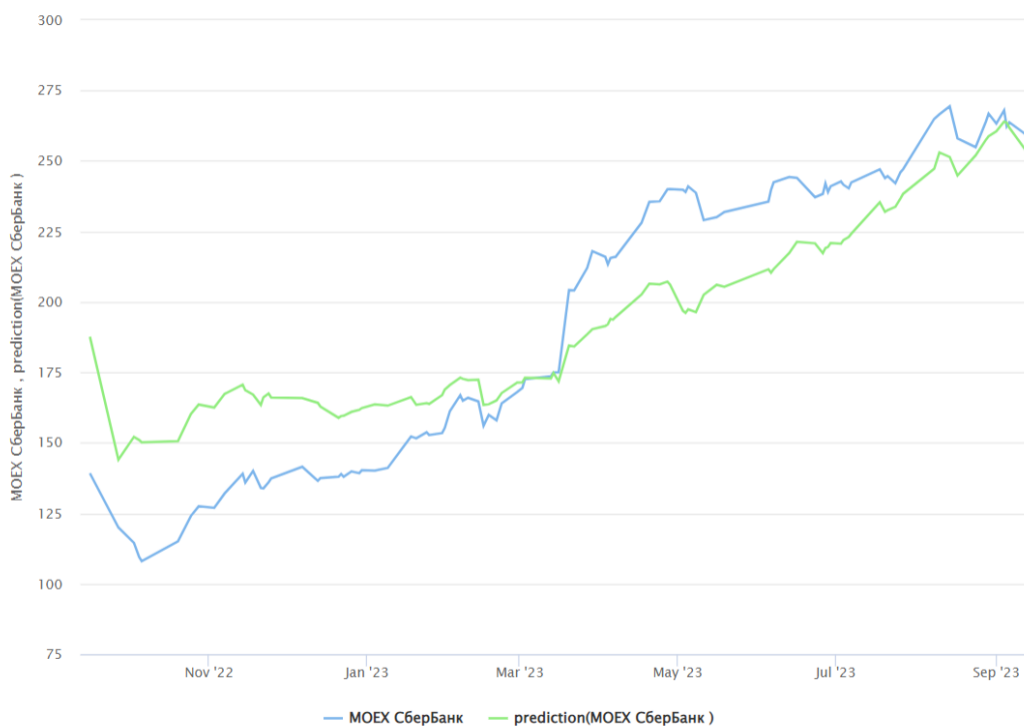


Рисунок 7 – Режим Visualizations

В данном случае можно сделать вывод о наличии умеренной положительной связи между переменными. Однако, стоит отметить, что связь между зависимой и независимой переменными в анализируемом наборе данных не очень сильная либо на нее могут влиять другие факторы, которые не были учтены в модели.

Вывод: в ходе выполнения лабораторной работы удалось получить начальные навыки работы с программным комплексом RapidMinerStudio и решить задачи построения модели регрессии, подготовленной разработчиками пакета RapidMiner, и для собственного набора данных. Таким образом, методы регрессии позволяют:

1. Оценить влияние независимых переменных на зависимую переменную.
2. Построить модель, которая предсказывает значения зависимой переменной на основе значений независимых переменных.
3. Оценить статистическую значимость коэффициентов модели и получить информацию о их вкладе в предсказание зависимой переменной.
4. Оценить точность модели и провести анализ ее качества, используя метрики оценки качества моделей регрессии.
5. Использовать модель для прогнозирования значения зависимой переменной для новых наблюдений, основываясь на значениях независимых переменных.

Контрольные вопросы

1) Как переводится с латинского слово «регрессия»?

Лат. regressio «обратное движение, возвращение»

2) Что такое регрессия?

Регрессия – это статистическая мера, которая пытается определить степень взаимосвязи между одной зависимой переменной (то есть атрибутом метки) и рядом других изменяющихся переменных, известных как независимые переменные (обычные атрибуты).

3) Линия регрессии это ...

Пусть имеется две переменные $X = (x_1, x_2, \dots, x_j, \dots, x_n)$ и $Y = (y_1, y_2, \dots, y_n)$. Математическое уравнение, которое оценивает линейную регрессию имеет вид

$$Y = a + bx.$$

Здесь $b_i x_i$ – значения одной из составляющих уравнения регрессии для конкретной точки и i – той переменной вектора независимой переменной. Если используется метод наименьших квадратов, то осуществляется подбор коэффициентов уравнения регрессии a и b так, чтобы для каждого X минимизировать квадрат разности

4) Гипотеза линейной регрессии - ...

При построении линейной регрессии проверяется нулевая гипотеза о том, что генеральный угловой коэффициент линии регрессии β равен нулю.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. RapidMiner. URL: <https://rapidminer.com/> (дата обращения 02.10.2023)