

5 ЛАБОРАТОРНАЯ РАБОТА «АНАЛИТИЧЕСКАЯ ПЛАТФОРМА DEDUCTOR. ПРОГНОЗИРОВАНИЕ С ПОМОЩЬЮ ЛИНЕЙНОЙ РЕГРЕССИИ»

5.1 Решение задачи прогнозирования с помощью линейной регрессии

Линейная регрессия предназначена для получения прогноза непрерывных числовых переменных. Линейная регрессия необходима тогда, когда предполагается, что зависимость между входными факторами и результатом линейная. Достоинством её можно назвать быстроту обработки входных данных и простоту интерпретации полученных результатов.

Описание линейной регрессии дано по ссылкам:
<https://wiki.loginom.ru/articles/linear-regression.html>,
<https://wiki.loginom.ru/articles/multiple-linear-regression.html>.

Рассмотрим применение линейной регрессии на примере данных по продажам, находящихся в файле Trade.txt. Будем строить прогноз с помощью линейной регрессии сразу после обработчика «Скользящее окно» (см. лабораторную работу 4).

Для построения линейной регрессии необходимо запустить Мастер обработки и выбрать в качестве обработки данных Линейную регрессию (рисунок 1).

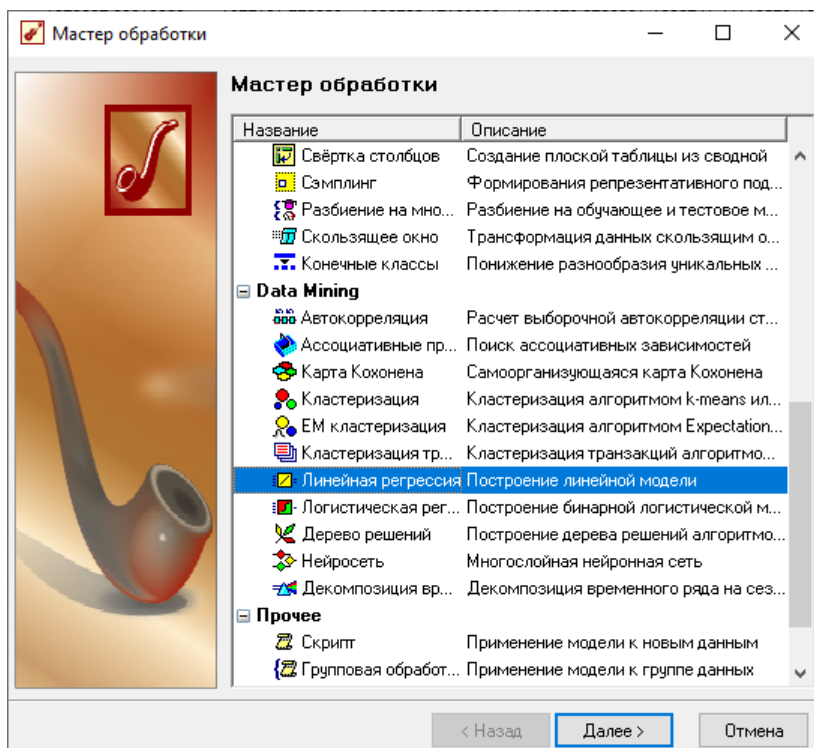


Рисунок 1 – Выбор узла «Линейная регрессия»

На первом шаге задаем назначение исходных столбцов. Предположим, что на прогноз влияет информация за 3 прошлых месяца, тогда укажем входными столбцами поля: «Количество – 3», «Количество – 2», и «Количество – 1». В качестве выходного поля укажем столбец «Количество». Остальные поля сделаем информационными (рисунок 2).

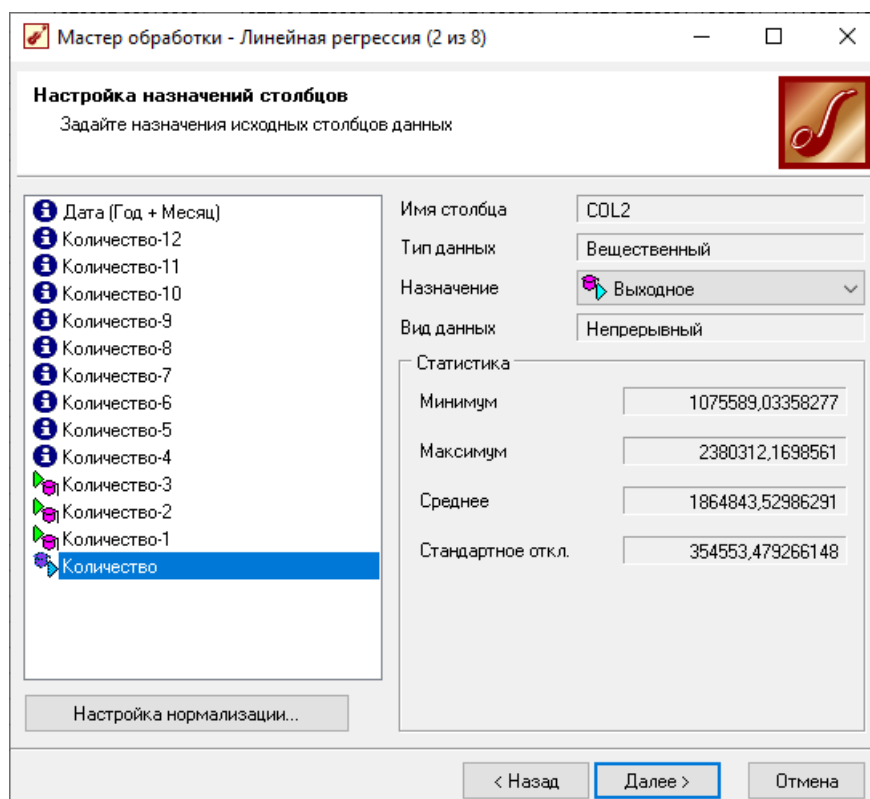


Рисунок 2 – Настройка назначений столбцов

На следующем шаге происходит настройка обучающего и тестового множеств, способ разложения исходного множества данных.

Третий шаг установки позволяет осуществить ограничение диапазона входных значений. Данный шаг оставим без изменений. При нажатии на кнопку «Далее» появляется окно запуска процесса обучения. В процессе выполнения видно, какая часть распознана на этапе обучения и теста.

После выполнения процесса выберем в качестве способа отображения визуализаторы «Коэффициенты регрессии», «Отчёт по регрессии», «Диаграмма рассеяния», «Что-если».

Например, диаграмма рассеяния приведена на рисунке 3. На рисунке 4 представлен фрагмент визуализатора «Отчёт по регрессии». Нужно обратить внимание на полученный коэффициент детерминации. Значение близко к 1, т.е. качество модели значительное.

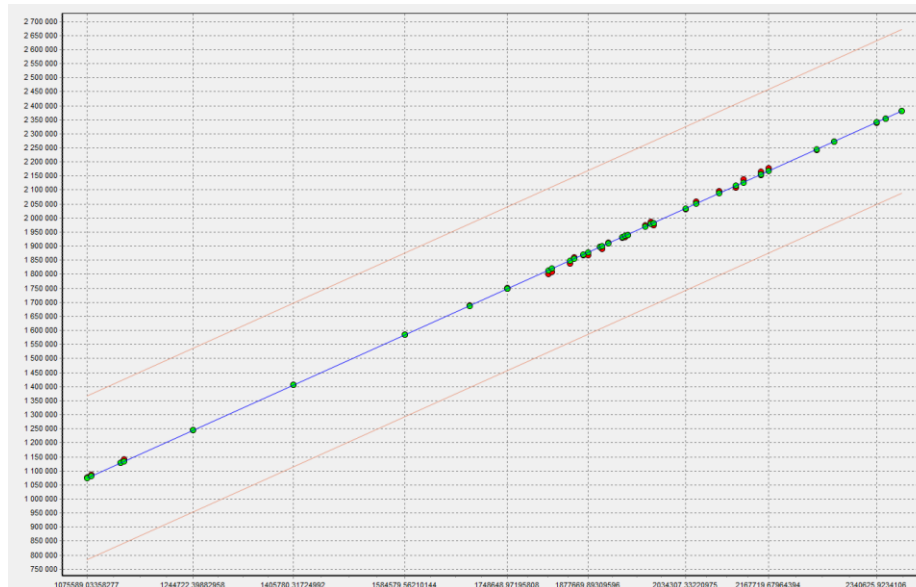


Рисунок 3 – Диаграмма рассеяния

Регрессия "1"					
Множ. коэффициент корреляции, R	Коэффициент детерминации, R ²	Скоррект. коэффициент детерминации	Стандартное отклонение	Размер выборки	Метод отбора переменных
0,9998	0,9997	0,9996	6782,9458	36	Полное включение

Рисунок 4 – Фрагмент визуализатора «Отчёт по регрессии»

Коэффициент детерминации изменяется в диапазоне от 0 до 1. Если он равен 0, это означает, что связь между переменными регрессионной модели отсутствует и вместо неё для оценки значения выходной переменной можно использовать простое среднее её наблюдаемых значений. Напротив, если коэффициент детерминации равен 1, это соответствует идеальной модели, когда все точки наблюдений лежат точно на линии регрессии, т.е. сумма квадратов их отклонений равна 0.

На практике, если коэффициент детерминации близок к 1, это указывает на то, что модель работает очень хорошо (имеет высокую значимость), а если к 0, то это означает низкую значимость модели, когда входная переменная плохо «объясняет» поведение выходной, т.е. линейная зависимость между ними отсутствует. Очевидно, что такая модель будет иметь низкую эффективность. Ссылка на материал о коэффициенте детерминации: <https://wiki.loginom.ru/articles/coefficient-of-determination.html>.

На рисунке 5 – визуализатор «Коэффициенты регрессии». Общий вид уравнения множественной регрессии: $Y=b_0+b_1x_1+b_2x_2+\dots+b_nx_n$ (материал в ранее

указанных ссылках). Полученное уравнение регрессии имеет вид:
 $Y=38698,00516+0,910493835x_1-2,592150129x_2+2,663537925x_3$.

Атрибут	Коэффициент
9.0 <Константа>	38698,00516
9.0 Количество-3	0,910493835
9.0 Количество-2	-2,592150129
9.0 Количество-1	2,663537925

Рисунок 5 – Визуализатор «Коэффициенты регрессии».

Теперь для построения прогноза запустим Мастера обработки, в котором выберем узел «Прогнозирование». На первом шаге обработчика происходит настройка связи столбцов для прогнозирования. Укажем связь между столбцами и горизонт прогноза равный 3.

На следующем шаге задаются параметры визуализации. Для данного примера выбираем отображение результатов в виде диаграммы прогноза. Теперь аналитик может дать прогноз о продажах, основываясь на модели, построенной с помощью линейной регрессии (рисунок 6).

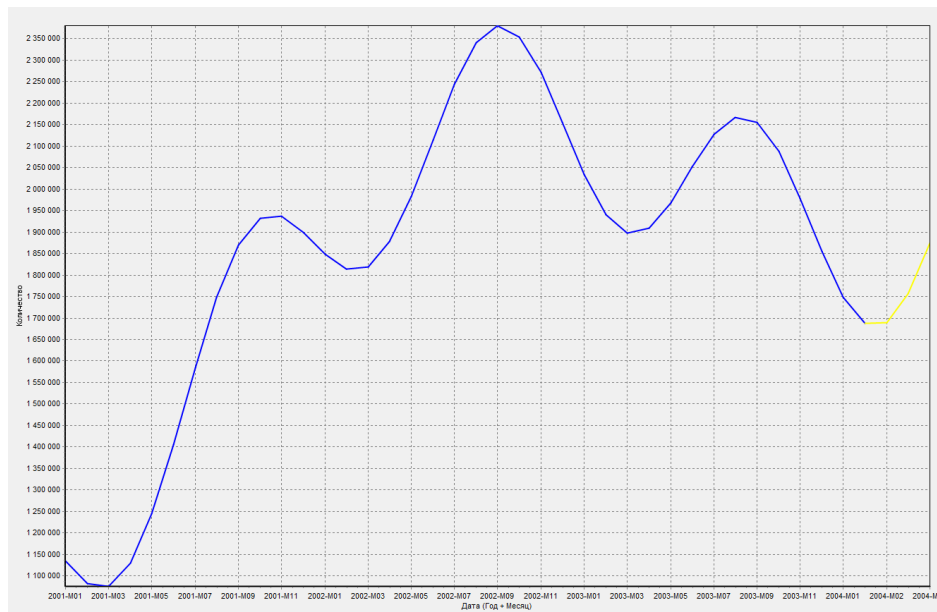


Рисунок 6 – Диаграмма прогноза

5.2 Задание и рекомендации

1. Изучить материал, представленный в лабораторной работе 5, а также в указанных ссылках.

2. Для решения задачи прогнозирования с помощью линейной регрессии использовать набор данных «Trade.txt».

3. Решить задачу прогнозирования с помощью линейной регрессии в Deductor Studio Academic по исходному набору данных, получить прогноз. В качестве входных столбцов задавать любые три поля. С помощью различных визуализаторов проанализировать полученные результаты, сделать выводы.

4. Подготовить отчёт и сдать преподавателю в электронной форме. Отчёт должен содержать краткое описание выполняемых действий, скриншоты, выводы. Наличие выводов является обязательным требованием.

5.3 Вопросы для защиты лабораторной работы

1. Для чего предназначена линейная регрессия?
2. Какой вид имеет множественное уравнение регрессии?
3. Что показывает коэффициент детерминации?
4. Для чего используются визуализаторы «Коэффициенты регрессии», «Отчёт по регрессии», «Диаграмма рассеяния»?
5. Для чего используются диаграмма прогноза?