

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

ПРЕПОДАВАТЕЛЬ

Доцент, к.т.н.

должность, уч. степень,
звание

подпись, дата

А. Д. Державина

инициалы, фамилия

ЛАБОРАТОРНАЯ РАБОТА № 4

по курсу: Интеллектуальные информационные системы

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ ГР.

№

8026



подпись, дата

М. В. Афанасьев

инициалы, фамилия

Санкт-Петербург

2023

Цель: изучить тему регрессии и применить полученные знания для создания модели предсказания параметров.

1 этап

В начале работы было инициирование выполнение модели линейной регрессии, подготовленной разработчиками пакета RM. Программная реализация модели линейной регрессии хранится в репозитории. Собранная программа показана на рисунке 1, часть исходных данных показана на рисунке 2, результаты на рисунке 3.

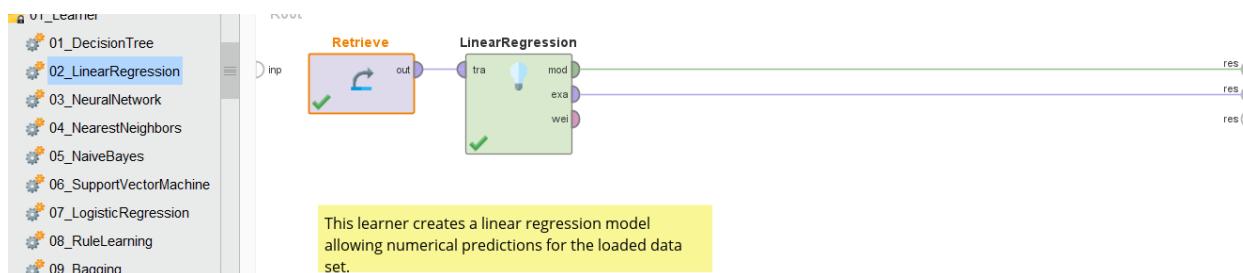


Рисунок 1 – Собранный программа

Row No.	label	a1	a2	a3	a4	a5
1	88.016	0.637	8.442	2.116	2.772	8.448
2	138.755	4.595	4.388	4.926	2.682	8.618
3	38.284	4.292	0.861	9.158	6.607	4.085
4	82.148	8.560	3.856	1.038	1.688	2.893
5	33.790	2.272	1.834	6.301	1.938	2.070
6	7.389	3.956	0.344	4.337	7.563	8.591
7	191.453	2.220	7.843	6.463	8.732	9.282
8	15.349	6.105	1.075	1.163	9.241	7.239
9	97.405	2.691	3.572	7.806	6.843	3.054
10	111.497	4.263	7.369	0.821	6.688	3.976
11	203.574	1.364	9.049	8.860	2.573	1.538
12	140.767	6.226	8.381	0.351	3.639	6.234
13	259.381	9.812	3.942	5.305	6.661	7.118
14	20.330	1.383	1.469	7.947	2.043	0.030

Рисунок 2 – Исходные данные

Attribute	Coefficient
a1	31.736
a2	42.948
a3	23.773
a4	3.706
a5	-4.184
(Intercept)	-304.228

Рисунок 3 – Результаты

На результатах видно влияние атрибутов a1, a2, ... a5 на атрибут label. Коэффициенты регрессии показывают величину и направление влияния независимых переменных на зависимую переменную в модели регрессии. Положительный коэффициент регрессии (a1-a4) указывает на прямую пропорциональность между независимой и зависимой переменными, тогда как отрицательный (a5) коэффициент регрессии указывает на обратную пропорциональность.

2 этап

Далее была сформирована модель регрессии для собственного набора данных. Выполнен ввод и необходимая подготовка данных. Выбран алгоритм регрессии – «Polynomial Regression» в RM.

Часть данных собственного набора представлена на рисунке 4. Данный набор представляет собой данные о красном вине, эти данные содержат метку label – quality (качество). Остальные столбцы описывают различные характеристики красного вина, например – хлориды и плотность.

Затем была построена модель регрессионного анализа, предварительно разделив исходные данные на обучающую и тестовую выборку, чтобы получить количественные показатели влияния различных характеристик вина на его качество. Схема модели на рисунке 5, на рисунке 6 результат.

Row No.	quality	volatile acidi...	citric acid	chlorides	free sulfur d...	total sulfur d...	density	sulphates
1	5	0.700	0	0.076	11	34	0.998	0.560
2	5	0.880	0	0.098	25	67	0.997	0.680
3	5	0.760	0.040	0.092	15	54	0.997	0.650
4	6	0.280	0.560	0.075	17	60	0.998	0.580
5	5	0.700	0	0.076	11	34	0.998	0.560
6	5	0.660	0	0.075	13	40	0.998	0.560
7	5	0.600	0.060	0.069	15	59	0.996	0.460
8	7	0.650	0	0.065	15	21	0.995	0.470
9	7	0.580	0.020	0.072	9	18	0.997	0.570

Рисунок 4 – Исходные данные

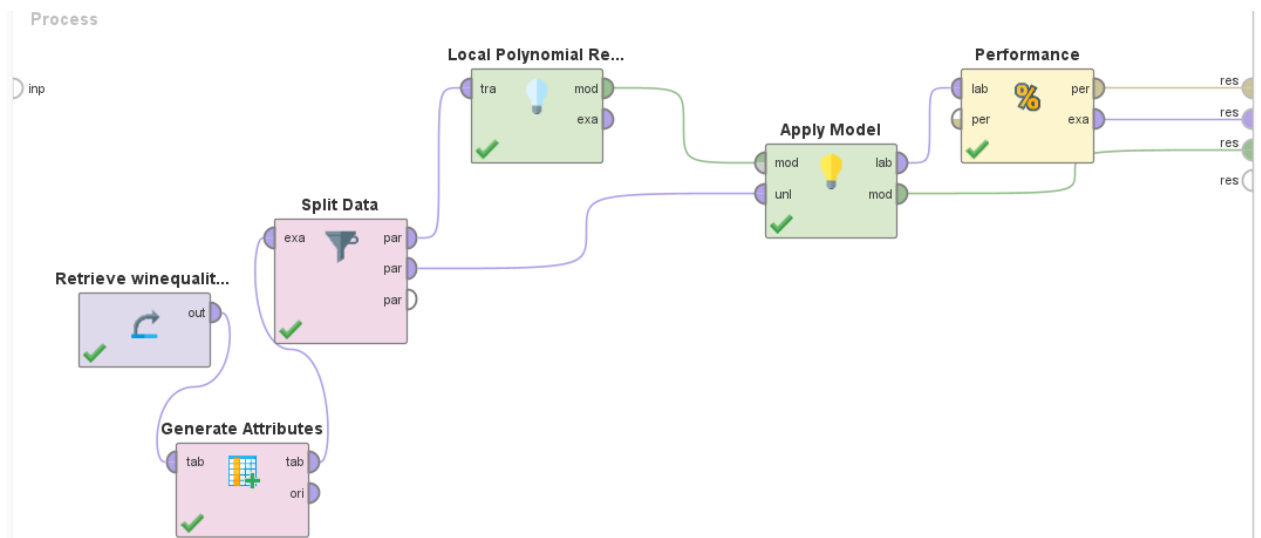


Рисунок 5 – Схема модели

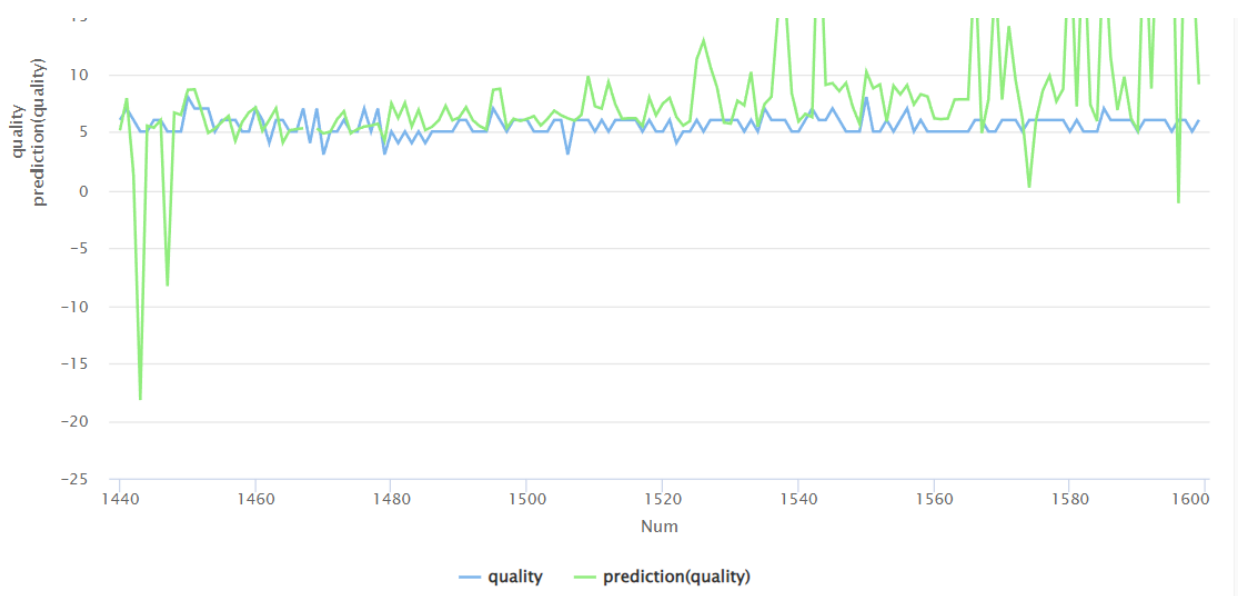


Рисунок 6 – Визуализированные результаты

Как видно из результатов модель получилась средней точности – среднее отклонение 5.9, при 8-ми бальной шкале оценивания, хотя многие предсказания близки находятся случаи, когда модель дает выходящие за пределы возможных значений результаты.

Изменив некоторые настройки аппарата регрессии и увеличив обучающую выборку, были получены заметно более точные результаты со средним отклонением 1.5, что почти в 4 раза лучше, по сравнению с первой попыткой. Результаты более точной модели показаны на рисунке 7.

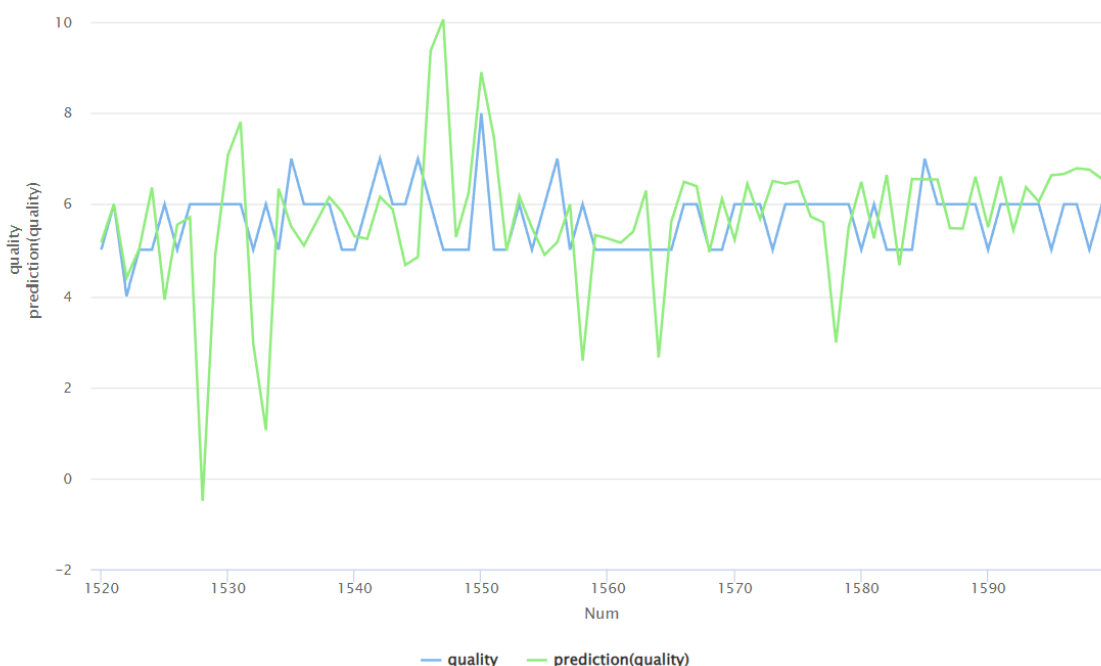


Рисунок 6 – Визуализированные результаты доработанной модели

Контрольные вопросы:

- 1) Как переводится с латинского слово «регрессия»? – 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 метод анализа данных, который предсказывает ценность неизвестных данных с помощью другого связанного и известного значения данных.

4) Гипотеза линейной регрессии - Гипотеза линейной регрессии предполагает, что зависимость между независимыми переменными (предикторами) и зависимой переменной может быть описана линейной функцией. То есть, предполагается, что существуют коэффициенты, которые связывают каждую независимую переменную с зависимой переменной с помощью линейного соотношения.

Вывод: в ходе выполнения лабораторной работы была изучена тема регрессии и регрессионного анализа. На основе этих данных была построена модель, которая была обучена за счет аппарата регрессии и смогла достаточно точно предсказать нужный параметр.