

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ
УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ (ТУСУР)
Кафедра экономической математики, информатики и статистики (ЭМИС)

ПРЕДОБРАБОТКА ДАННЫХ

Отчет по практической работе по дисциплине «Информационно-
аналитические системы управления»

Студент гр. 590-1

_____/П.А. Отегов

«__» _____ 2023 г.

Ассистент кафедры ЭМИС

_____/ К.В. Никитин

оценка подпись

«__» _____ 2023 г.

Томск 2023

«Предобработка данных»

Цель работы: получить навыки предобработки данных в АП Deductor.

Задание.

1 Сгенерировать условные исходные данные, преднамеренно снизив их полноту и структурную чистоту, используя функцию $y = 1/0.5^x$ и отрезок x в промежутке $[0; 3]$. Импортировать данные в Deductor;

2 Произвести предобработку данных с целью повышения качества дальнейшего анализа.

Результат выполнения заданий

Пример выполнения задания 1 представлен на рисунках 1-2.

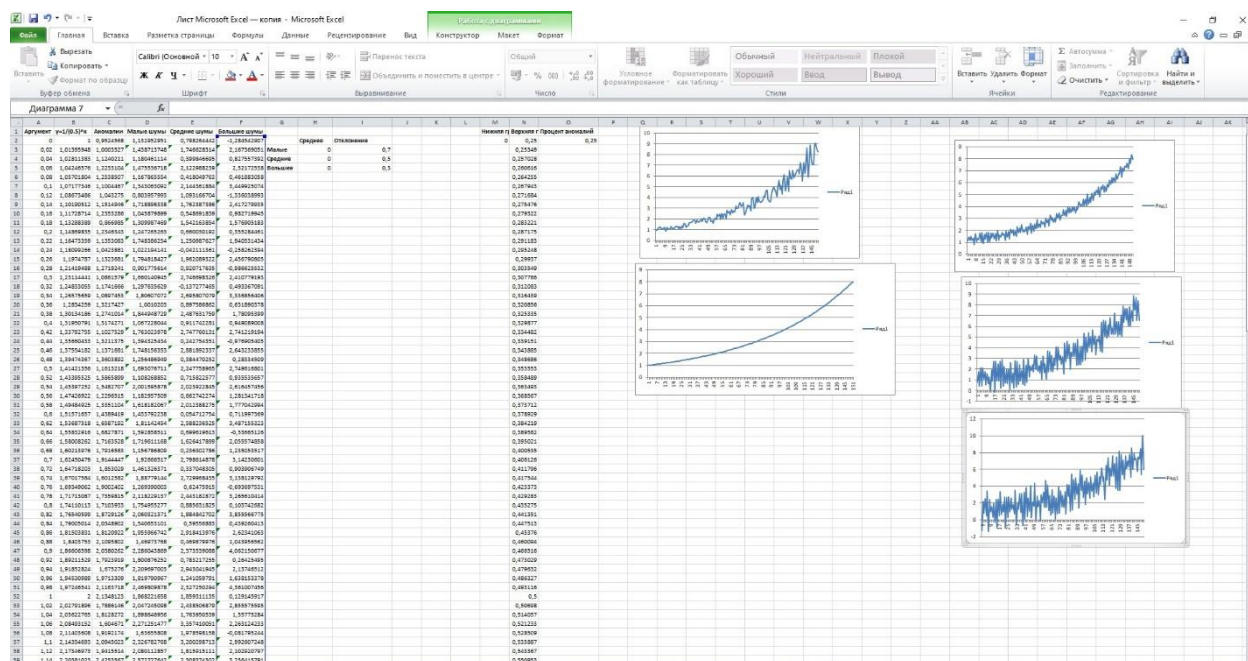


Рисунок 1 – Условные исходные данные со сниженной полнотой

Deductor Studio Academic (Новый) - [Настройка набора данных (-COL7, -COL8, -COL9, -COL10, -COL11, -COL12, -Нижняя гр, -Верхняя гр, -Процент аномалий)]

Файл Правка Вид Избранное Сервис Окно ?

Сценарии ? X

Сценарии

- Текстовый файл [C:\Users\540ps\Desktop\data.txt]
- Настройка набора данных (-COL7, -COL8, -COL9, -COL10, -COL11, -COL12, -Нижняя гр, -Верхняя гр, -Процент аномалий)

Таблица

Аргумент	u=1/0.5 ^{1/4}	Аномалии	Малые шумы	Средние шумы	Большие шумы
0	1	0.854970512	0.75256772	0.736720313	0.769118434
0.02	1.01395948	1.175287087	1.36327668	2.511719332	1.770114304
0.04	0	0.255473597	0.738949755	-0.485472185	
0.06	1.042465761	1.023992295	1.291189033	1.936113414	3.186382323
0.08	1.057018041	1.056799131	1.114585215	0.466256454	0.626870157
0.1	1.071773463	0.977385736	1.350976389	2.063678986	1.347436694
0.12	1.086734863	1.132367618	1.037284153	-0.160843952	-0.818570103
0.14	1.101905116	1.208715284	1.533313524	1.589959321	1.615801785
0.16	1.117287138	1.106125536	1.236547918	0.417200536	0.205379457
0.18	1.132883885	1.069137356	1.511848786	2.171811053	2.068443016
0.2	1.148693955	1.102841855	0.941544322	0.820545644	-0.607927109
0.22	1.164733586	1.310752515	1.310988273	1.462976277	3.225026577
0.24	1.180932661	1.059107832	1.052895992	1.044395653	1.160979222
0.26	1.197478705	1.388149376	1.575867654	1.62017388	3.139681678
0.28	0	0.176871745	-0.620009972	0.747813086	
0.3	1.231144413	1.381194813	1.426308195	2.754505929	2.137739956
0.32	1.248330549	1.28523699	1.105199507	0.32004367	1.127353312
0.34	1.265756594	1.378257226	1.59161407	1.779132109	1.596774447
0.36	1.283425898	1.164246023	0.960339723	0.246745386	1.018984431
0.38	1.301341855	1.116493494	1.538357105	2.313014095	3.24281615
0.4	1.319507911	1.400353599	0.936312766	0.532720873	1.031855103
0.42	1.337927555	1.47444	1.890621232	2.298154088	2.212988321
0.44	1.356604327	1.170534626	0.977295882	0.834765153	1.091157517
0.46	0	0.628794331	1.472535546	2.1361223	
0.48	1.394743666	1.393397269	1.306606905	0.861918259	0.519466362
0.5	0	0.618763485	1.906101778	2.919915721	
0.52	1.433955248	1.607999023	1.365602032	1.062862996	0.330963058
0.54	1.453972517	1.493040264	1.5838383	2.753343849	1.957018126
0.56	1.474269217	1.343358636	1.281602443	0.554177737	0.348948909
0.58	1.494843249	1.594972277	2.008551801	2.065363419	2.023532955
0.6	1.515715567	1.478388044	1.347029638	0.467651812	-0.459915321
0.62	1.536875181	1.511103573	1.795609267	1.58918724	1.876181525
0.64	1.558329159	1.528657006	1.106356749	0.190276327	1.183442264
0.66	1.580082624	1.256318281	1.907967911	2.745686467	3.305319859
0.68	1.602139755	1.377831691	1.155536142	0.59233589	0.668968808
0.7	1.624504793	1.50759497	1.902918893	2.834362207	2.725151791
0.72	0	0.413849851	-0.003691041	0.027166222	
0.74	1.670175839	1.459860245	1.715520868	2.408962522	2.027728428
0.76	1.634390625	1.631196087	1.427909599	1.28456264	0.36023497
0.78	1.717130873	1.963138872	1.304777291	1.779330436	1.831953897
0.8	1.741101127	1.607125462	1.297322034	0.528455842	-0.722632456
0.82	1.765405993	1.792842202	2.22054158	3.134905674	2.037815637
0.84	1.790050142	2.014501737	1.4006168	0.425626781	0.209443496
0.86	1.815038311	1.735202019	1.89276092	2.209099483	2.418673747
0.88	1.840375301	1.776830295	1.384949172	1.292329598	1.581929505
0.9	1.866065983	1.461086878	1.899199333	2.954795962	2.267253561
0.92	1.892115293	1.889264898	1.570443652	1.760471973	1.685626504
0.94	1.918528239	1.640205905	2.305884182	3.259667027	2.273510395
0.96	1.945309895	1.956718977	1.688893045	1.860633764	-0.277677944
0.98	1.972465409	2.053944516	2.119438136	2.413987016	3.078436709
1	2	1.672329382	1.896553493	1.826582567	1.221006294
1.02	2.02791896	1.852781063	2.478815539	3.319186068	3.303464509
1.04	2.056227653	1.676209099	1.967507331	1.999455442	1.10742034
1.06	2.084931522	1.833810181	2.449875406	2.373028344	3.330177141

Рисунок 2 – Импортированные в Deductor данные

Пример выполнения задания 2 представлен ниже.

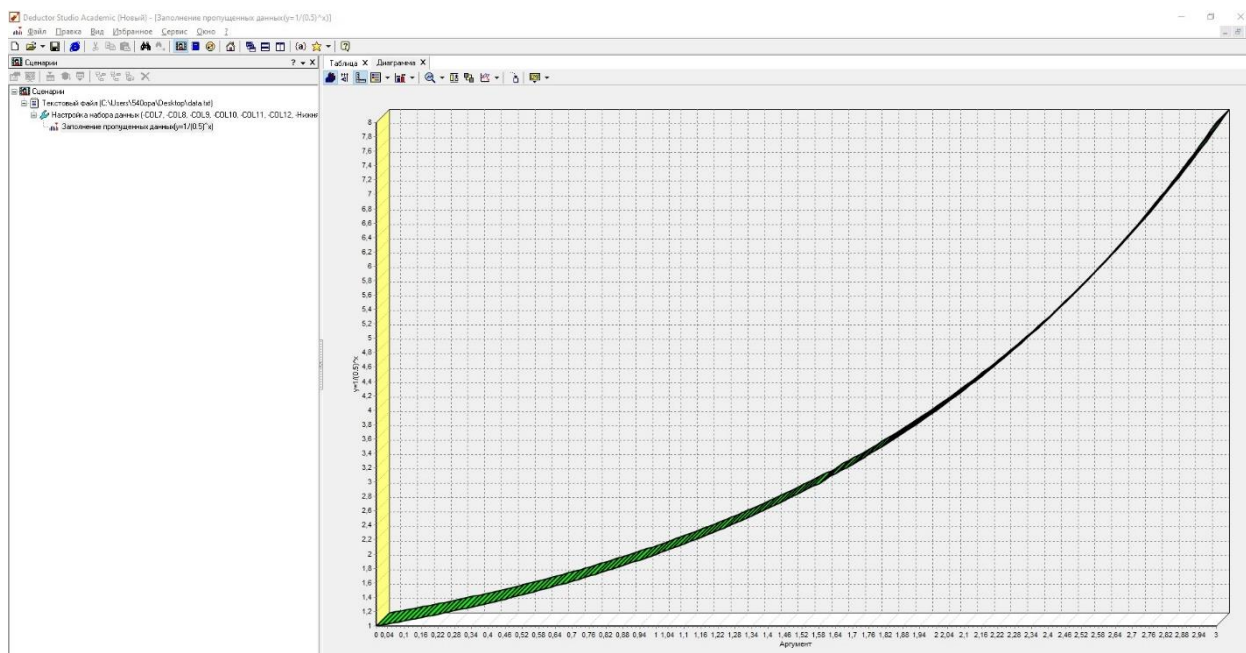


Рисунок 3 – Восстановленные с помощью инструмента «Заполнение пропущенных данных» данные

Далее для восстановления была выбрана колонка с аномальными данными. Для её восстановления был использован инструмент «Спектральная обработка» с параметрами, представленными на рисунке 4.

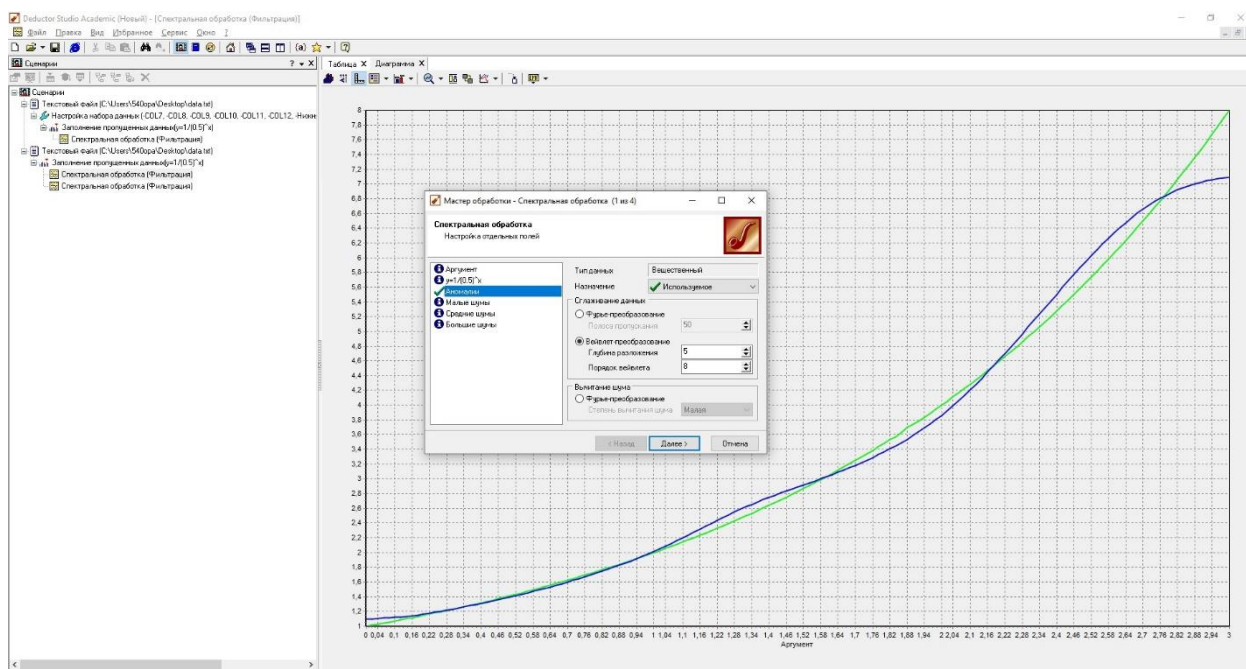


Рисунок 4 – Восстановленные аномальные данные

Далее для восстановления были выбраны колонки с малыми, средними и большими шумами. Для её восстановления был использован инструмент «Спектральная обработка» с параметрами, представленными на рисунках 5-7 для малых, средних и больших шумов соответственно.

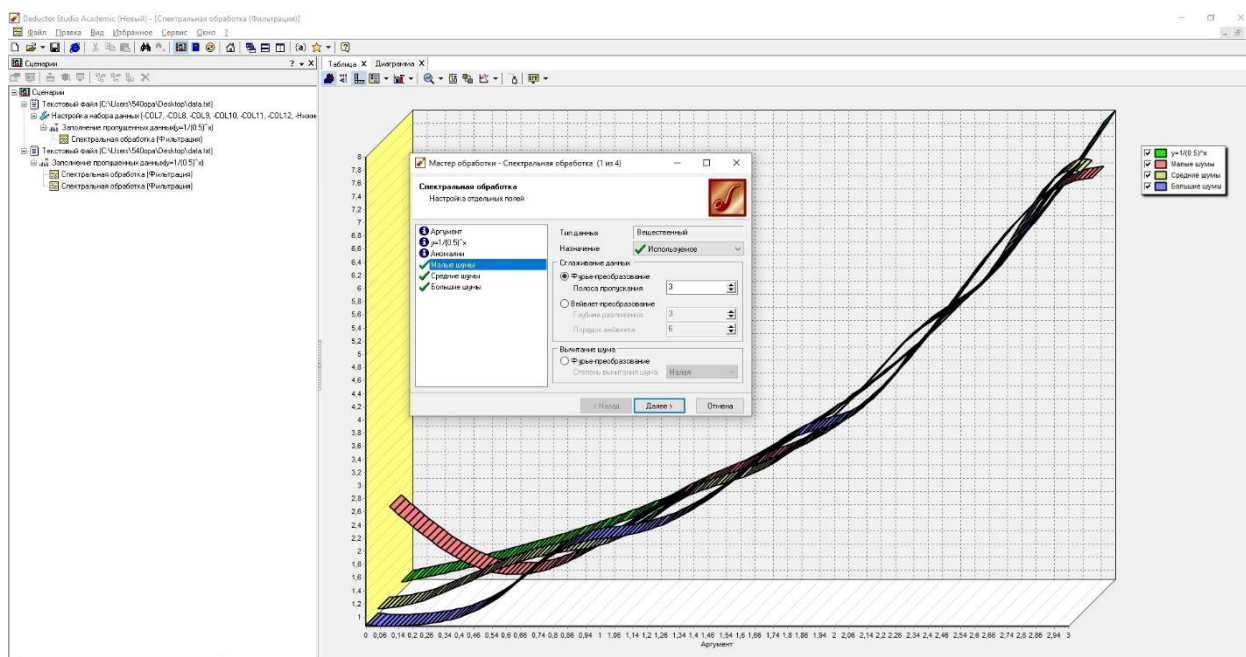


Рисунок 5 – Восстановленные малые шумы

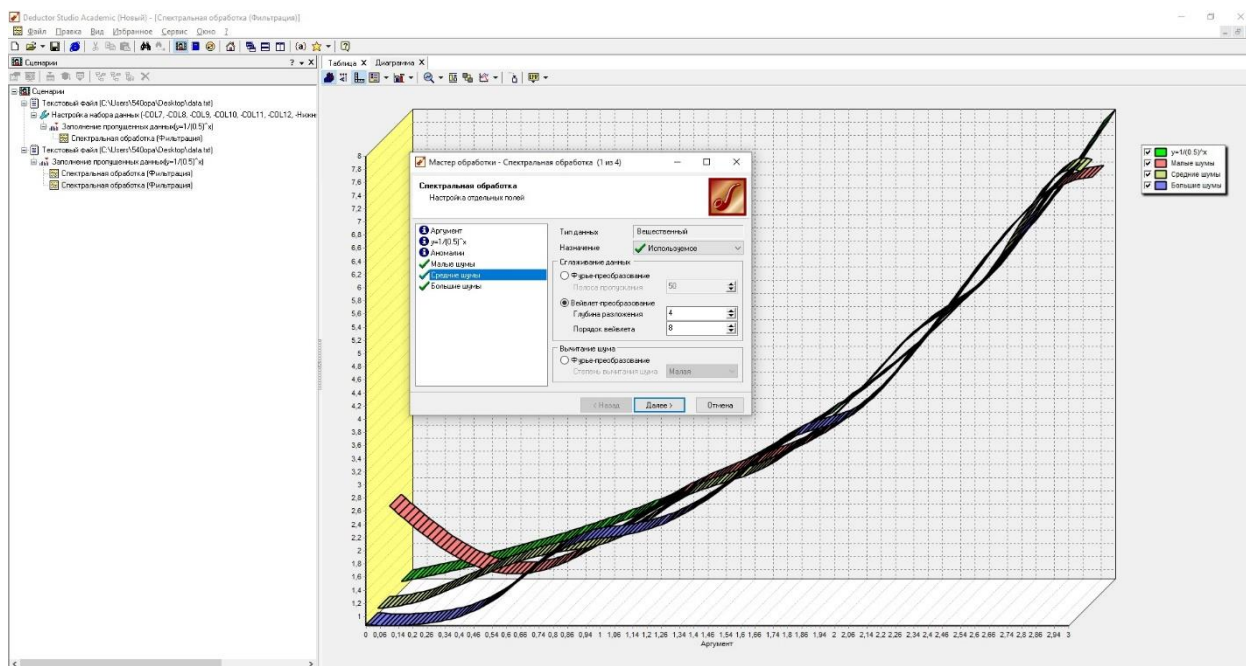


Рисунок 6 – Восстановленные средние шумы

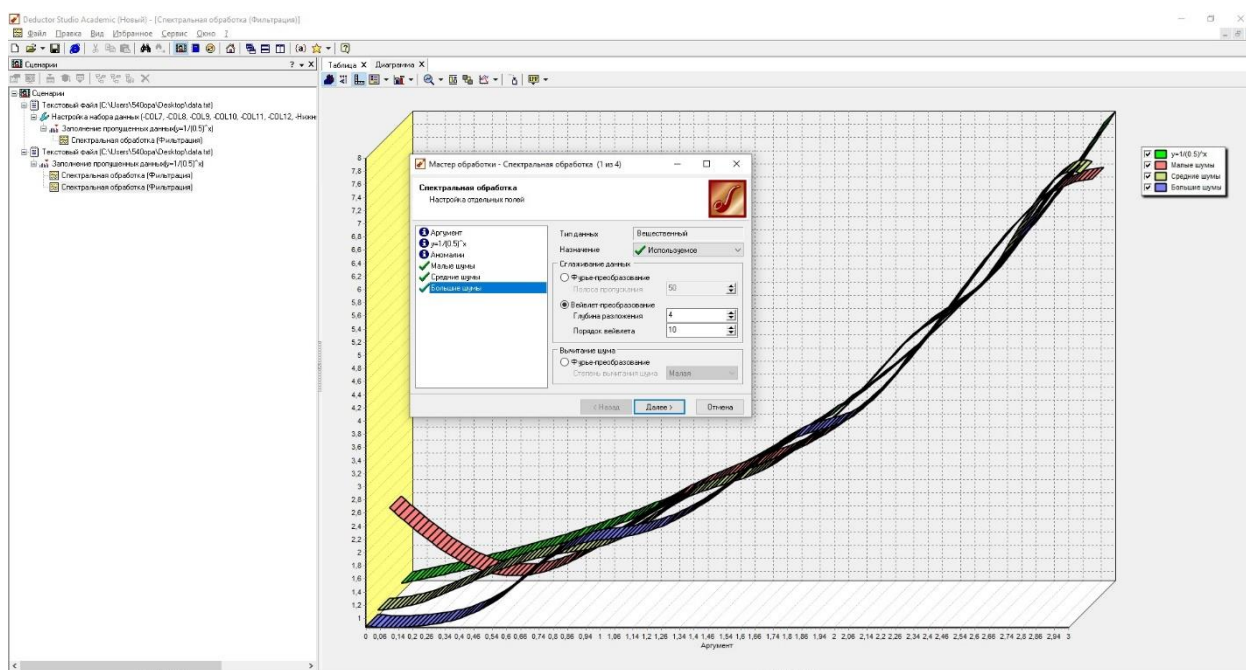


Рисунок 7 – Восстановленные большие шумы

Вышеописанные параметры наиболее корректно восстанавливают данные, что было проверено эмпирически. Для удостоверения в этом был использован инструмент «Качество данных», который показал пригодность данных – это отображено на рисунке 8.

The screenshot shows the 'Оценка качества данных' (Data Quality Assessment) window in Deductor Studio Academic. The left sidebar lists a workflow: 'Текстовый файл (C:\Users\Babara\OneDrive\data\...)' -> 'Настройка набора данных (COL1, COL8, COL9, COL10, COL11, COL12, ...)' -> 'Заполнение пропущенных данных (1/0/5)' -> 'Спектральная обработка (Фильтрация)' -> 'Спектральная обработка (Фильтрация)' -> 'Оценка качества данных (1/0/5)'. The main table displays the results of the quality assessment for five data columns.

№	Столбец	Тип данных	Вид данных	Пропуски		Выбросы		Экстремальные		Качество данных	Результат
				Колво	Доля	Колво	Доля	Колво	Доля		
1	y=1/0/5/x	8.8 Вещественный	Непрерывный							0.9241	Пригоден
2	Амплитуда	8.8 Вещественный	Непрерывный							0.8778	Пригоден
3	Малые шумы	8.8 Вещественный	Непрерывный							0.9112	Пригоден
4	Средние шумы	8.8 Вещественный	Непрерывный							0.9434	Пригоден
5	Большие шумы	8.8 Вещественный	Непрерывный							0.9427	Пригоден

Рисунок 8 – Пригодные данные

Вывод: в процессе выполнения работы были получены навыки предобработки данных в АП Deductor