

Министерство науки и высшего образования Российской Федерации

Севастопольский государственный университет

Кафедра ИС

ОТЧЕТ

Расчетно-графическая задание

«Компьютерные методы анализа данных и прогнозирования»

по дисциплине

«ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ»

Выполнил студент группы ИС/б-17-2-о

Горбенко К.Н.

Проверил

Сырых О.А.

Севастополь

2020

1 ЦЕЛЬ РАБОТЫ

- Приобрести базовые навыки работы в Deductor Studio.
- Изучить основы методов анализа экспериментальных данных и освоить технику их практического применения в Deductor Studio.

2 ПОСТАНОВКА ЗАДАЧИ

Выполнить 4 задания в программе дедуктор, следуя методическим указаниям. Подобрать данные и провести поиск ассоциативных правил и прогнозирование временного ряда.

3 ИЗУЧЕНИЕ DEDUCTOR

1. Создать проект в дедукторе. Изменить его свойства и открыть в любом текстовом редакторе. Сделать видимой вкладку Подключения.

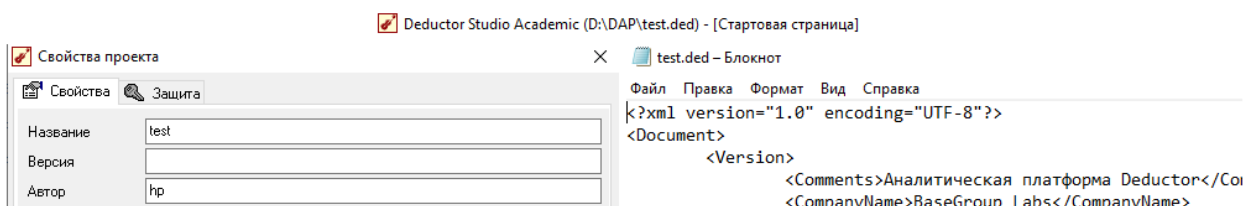


Рисунок 3.1 – Выполнение задания номер 1

2. Создать новый проект. Создать текстовый файл и импортировать его в дедуктор. После исправления в программе экспортировать результат в текстовый файл и отобразить результат.

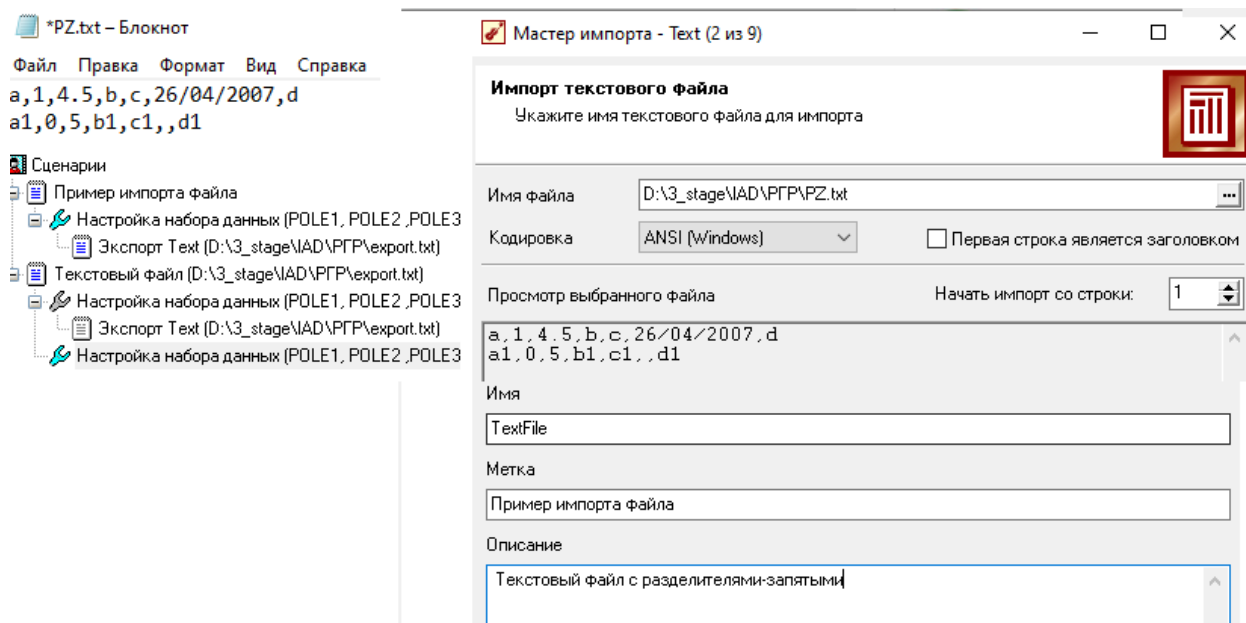


Рисунок 3.2 – Выполнение задания номер 2

3. В проекте из предыдущего задания настроить визуализаторы. В визуализаторе таблицы настроить, чтобы при отображении поля 3 добавлялось слово кг



Рисунок 3.3 – Выполнение задания номер 3

4. Ответы на вопросы

4.1. Дедуктор состоит из Warehouse, Studio, Viewer, Server, Client

4.2. Чтобы скрыть столбец из набора данных, нужно задать ему назначение Неиспользуемое.

4.3. Категории пользователей аналитик; – пользователь; – администратор; – программист

4.4. Функции аналитика создание в Deductor Studio сценариев – последовательности шагов, которую необходимо провести для получения нужного результата. – построение, оценка и интерпретация моделей. – настройка панели отчетов для пользователей Deductor Viewer. – настройка сценария на поточную обработку новых данных

4.5. В deductor studio ключевым понятием является проект. Это файл с расширением*.ded, по структуре соответствующий стандартному xml-файлу. Он хранит в себе: – последовательности обработки данных (сценарии); – настроенные визуализаторы; – переменные проекта и служебную информацию.

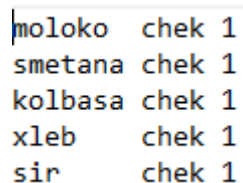
5. Выводы

В ходе лабораторной работы изучены основные возможности и компоненты программы Deductor Studio. Изучены основные функции аналитика, пользователя, администратора, программиста в программе Deductor Studio.

4 ПОИСК АССОЦИАТИВНЫХ ПРАВИЛ

1. Подготовка данных.

В качестве исходных данных для анализа возьмем 30 чеков покупателей



```
moloko chek 1
smetana chek 1
kolbasa chek 1
xleb chek 1
sir chek 1
```

Рисунок 4.1 – Исходные данные

2. Сформируем текущие данные в таблицу, состоящую из двух столбцов:

Продукт и Номер чека и укажем идентификатор и элемент транзакции.

Символом-разделителем является
☒ Символ табуляции ☐ Пробел ☐ Точка
☐ Точка с запятой ☐ Запятая ☐ Другой
☐ Считать последовательные разделители одним

COL1	COL2
moloko	chek 1
smetana	chek 1
kolbasa	chek 1
xleb	chek 1
sir	chek 1
min.voda	chek 2
sigareti	chek 2

ab Товар

ab Номер чека

Имя столбца: COL2
Метка столбца: Номер чека
Тип данных: Строковый
Вид данных: Дискретный
Назначение: ID Транзакция

Рисунок 4.2 – Определение столбца идентификатора транзакции и ее элемента

3. Для поиска ассоциативных правил воспользуемся Мастером обработки, где выберем тип обработки “Ассоциативные правила”.

Мастер обработки - Ассоциативные правила (3 из 6)
Настройка параметров построения ассоциативных правил
 Укажите значения параметров построения ассоциативных правил

Часто встречающиеся множества
Минимальная поддержка, %: 13
Максимальная поддержка, %: 80
☐ Максимальная мощность искомым часто встречающихся множеств: 4

Ассоциативные правила
Минимальная достоверность, %: 60
Максимальная достоверность, %: 90

< Назад

Далее >

Отмена

Рисунок 4.3 –Настройка параметров построения правил

Последующие действия позволяет запустить процесс поиска ассоциативных правил.



Рисунок 4.4 – Результат поиска ассоциативных правил

4. Популярные наборы

№	Номер множества	Элементы	Поддержка		Мощность
			Кол-во	%	
95	54	sodovaya vipechka	4	13,33	2
96	55	sodovaya wino	4	13,33	2
97	56	sodovaya xleb	9	30,00	2

Рисунок 4.5 – Популярные наборы

Исследуя визуализатор поиска ассоциативных правил “Популярные наборы”, можно сделать вывод, что такие продукты, как: колбаса, хлеб, минеральная вода являются приоритетными к покупке в торговой точке.

5 Визуализатор "Правила"

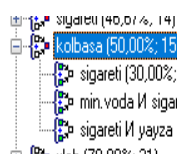
№	Номер правила	Условие	Следствие	Поддержка		Достоверность	Лифт
				Кол-во	%		
1	1	kolbasa	sigareti	9	30,00	60,00	1,286
2	2	sigareti	kolbasa	9	30,00	64,29	1,286
3	3	kolbasa	xleb	9	30,00	60,00	0,857
4	4	kolbasa	yayza	9	30,00	60,00	1,000

Рисунок 4.6 – Визуализатор Правила

Из полученных результатов на рисунке 7 видно, что при покупке конфет, покупатель с вероятностью 80% купит и яйца и содовую, при покупке Колбасы он купит Сигареты с вероятностью 60%.

6. Визуализатор “Дерево правил”.

При построении дерева правил по следствию на первом (верхнем) уровне находятся узлы со следствиями, а на втором уровне – узлы с условиями.

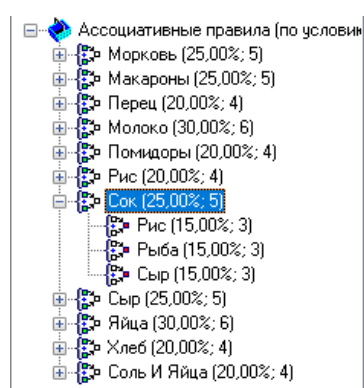


Условие	Поддержка		Достоверность, %	Лифт
	Кол-во	%		
sigareti (40,00%; 14)				
kolbasa (50,00%; 15)				
sigareti (30,00%; 9)				
min.voda И sigareti (13,33%; 4)				
sigareti И yayza (20,00%; 6)				
vish (70,00%; 21)				

Рисунок 4.7 – Построение дерева правил по следствию

Например, для того чтобы человек приобрел Колбасу, он должен купить хотя бы один предмет или пару предметов из списка: сигареты, минеральную воду и сигареты, сигареты и яйца.

Второй вариант дерева правил – дерево, построенное по условию. Здесь на первом уровне располагаются узлы с условием.



Количество правил: 3; Условие: Sok				
Следствие	Поддержка		Достоверность, %	Лифт
	Кол-во	%		
Рис	3	15,00	60,00	3
Рыба	3	15,00	60,00	2
Сыр	3	15,00	60,00	2,4

Рисунок 4.8 – Построение дерева правил по условию

Узлы - верхний уровень дерева и условие. А ветви – следствия. Это означает, что покупатель, купивший колбасу, так же купит сигареты, хлеб и яйца с достоверностью 60%.

7 Анализ “Что-если” позволяет определить, что получим в качестве следствия, если выберем определенные условия. Например, какие товары приобретаются совместно с выбранными товарами. Пусть необходимо

проанализировать, что, возможно, забыл покупатель приобрести, если он уже взял Сигареты и Сметану.

Условие	
Элемент	Поддержка, %
sigareti	46,67
smetana	40,00
<div><div></div><div></div></div>	
<div><div><div><div></div><div></div><div></div></div><div><div></div><div></div><div></div></div></div><div><div></div><div></div><div></div></div><div><div></div><div></div><div></div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div></div>	

Рисунок 4.9 – Результат анализа Что-Если

В данном случае появится Минеральная вода, Хлеб, Яйца и Колбаса. Именно эти продукты покупатель возможно забыл приобрести с вероятностью 65-75%.

5 ПРОГНОЗИРОВАНИЕ ВРЕМЕННОГО РЯДА

1. Для анализа временного ряда был выбрано количество осадков за 3 года, где COL1 – дата, COL2 –осадки, мм.

январь.18	69
фев.18	50
мар.18	51
апр.18	53
май.18	63
июн.18	74
июл.18	57
авг.18	55

Рисунок 5.1 – Фрагмент данных для анализа временного ряда

2. После импорта данных воспользуемся диаграммой для их просмотра и проведем удаление аномалий и сглаживание данных.

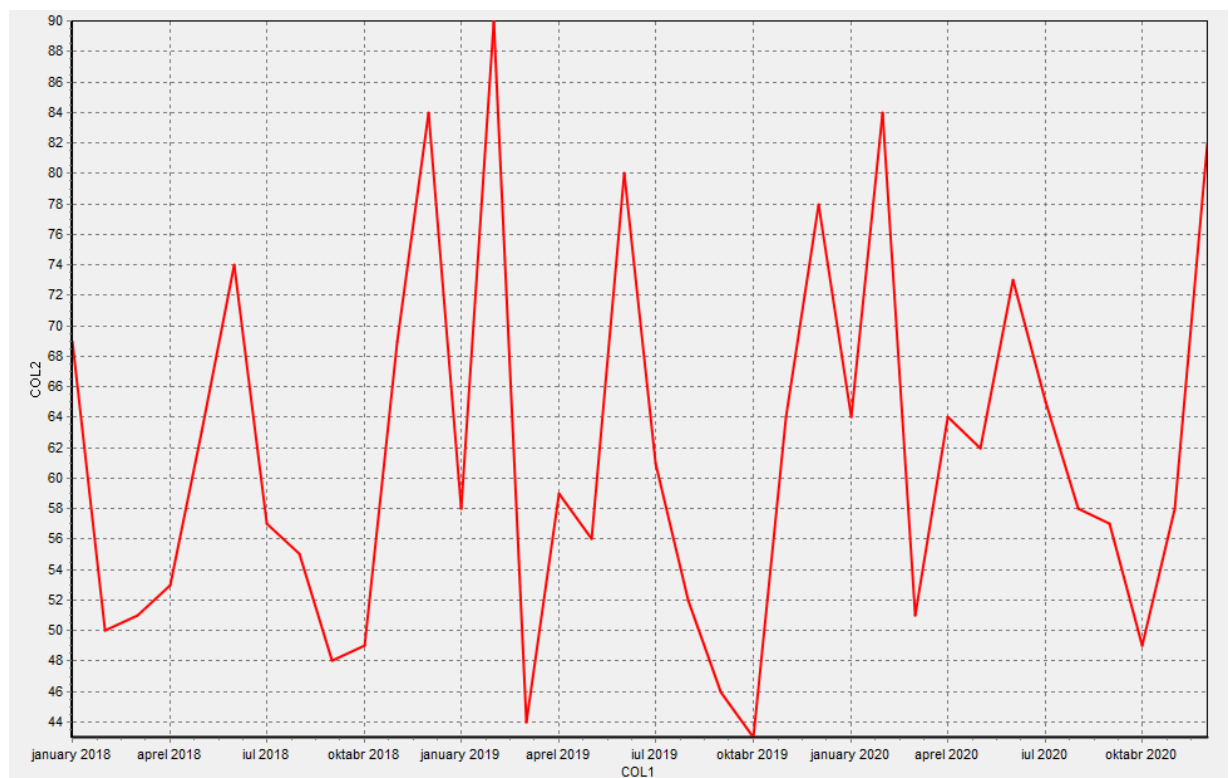


Рисунок 5.2 – Диаграмма до обработки

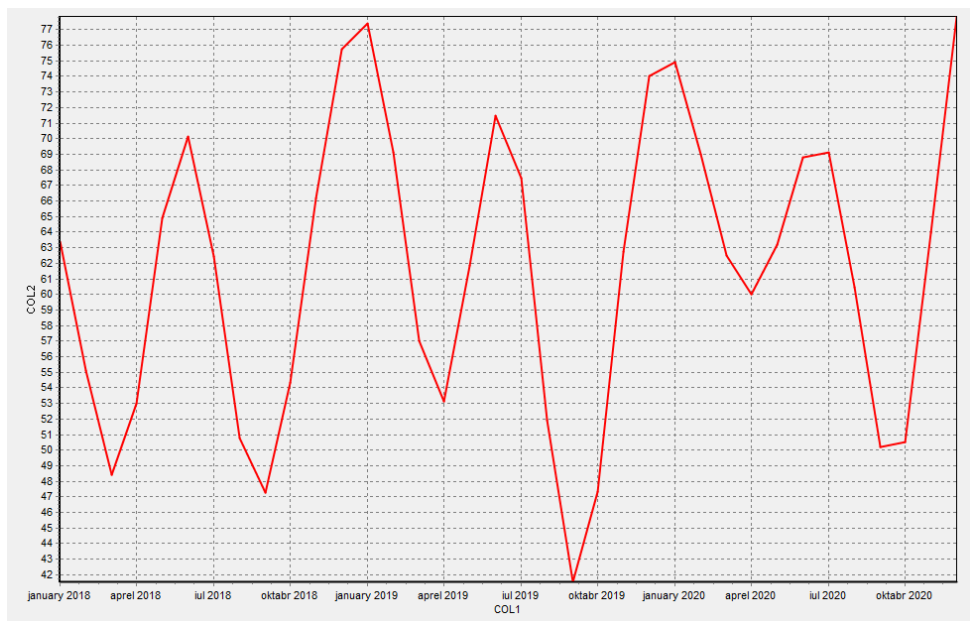


Рисунок 5.3 – Диаграмма после обработки

3. Отберем данные используя метод “Скользящее окно” с глубиной погружения 12 месяцев.

COL1	COL2-12	COL2-11	COL2-10	COL2-9	COL2-8	COL2-7	COL2-6	COL2-5	COL2-4	COL2-3
january 2019	69	50	51	53	63	74	57	55	48	49
febr 2019	50	51	53	63	74	57	55	48	49	69
mart 2019	51	53	63	74	57	55	48	49	69	84
aprel 2019	53	63	74	57	55	48	49	69	84	58
may 2019	63	74	57	55	48	49	69	84	58	90
iun 2019	74	57	55	48	49	69	84	58	90	44
iul 2019	57	55	48	49	69	84	58	90	44	59
avgust 2019	55	48	49	69	84	58	90	44	59	56
sentabr 2019	48	49	69	84	58	90	44	59	56	80
oktabr 2019	49	69	84	58	90	44	59	56	80	61
noyabr 2019	69	84	58	90	44	59	56	80	61	52
dekabr 2019	84	58	90	44	59	56	80	61	52	46

Рисунок 5.4 – Часть данных после использования метода “Скользящее окно”

4. Перейдем к обучению нейросети для дальнейшего построения прогноза.

В качестве входных полей будут COL2-1, COL2-2, COL2-3, COL2-4, COL2-5 а в качестве выходного COL2.

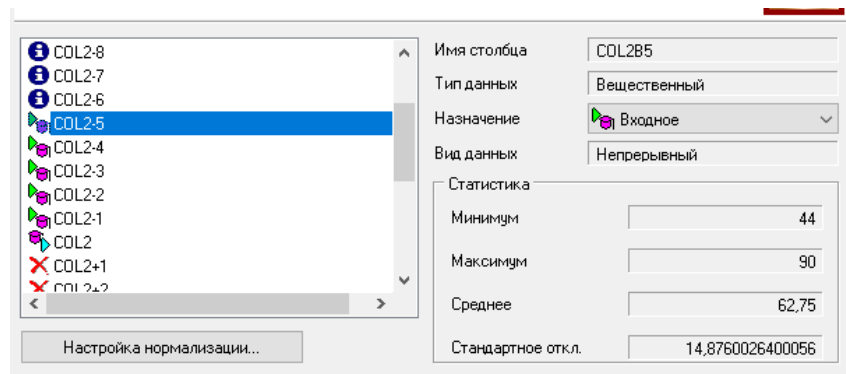


Рисунок 5.5 – Настройка назначений столбцов

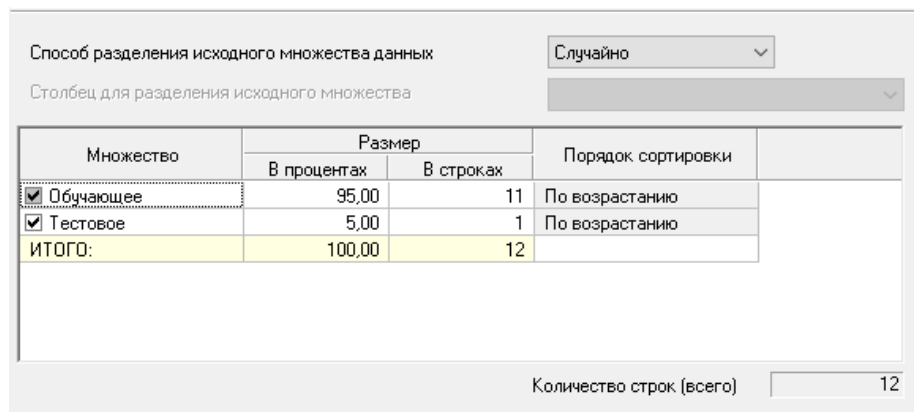


Рисунок 5.6 – Разбиение исходного набора данных на подмножества

Результатом обучения нейросети будет две диаграммы:

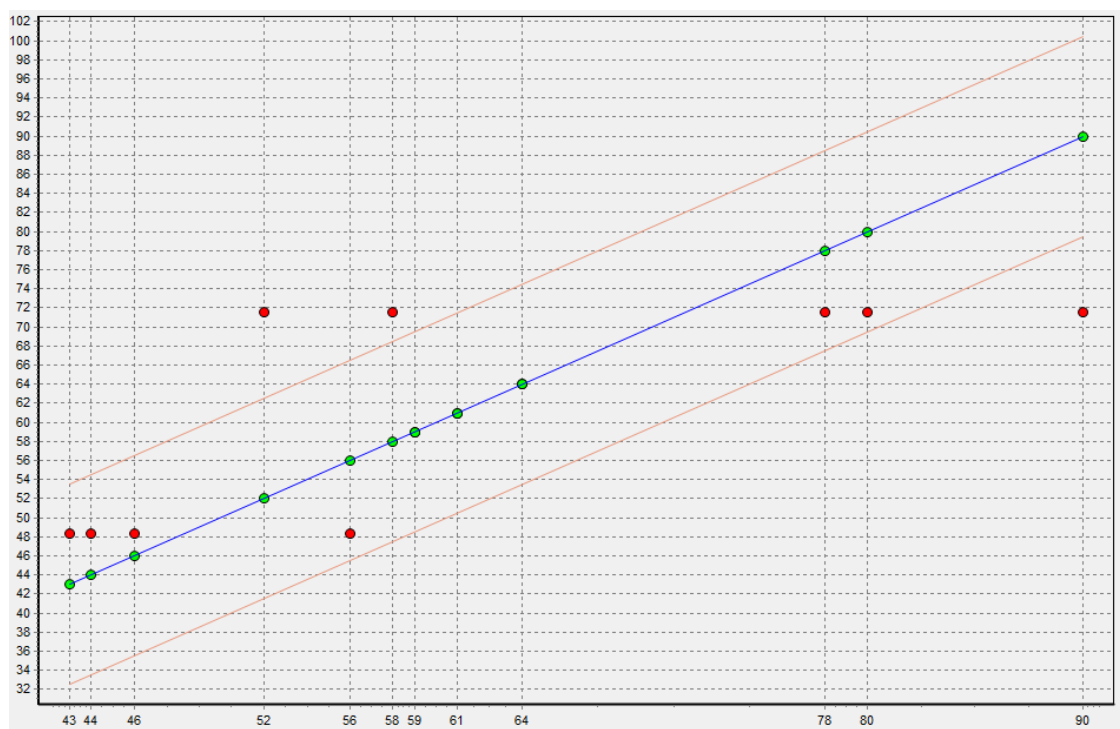


Рисунок 5.7 – Диаграмма рассеяния

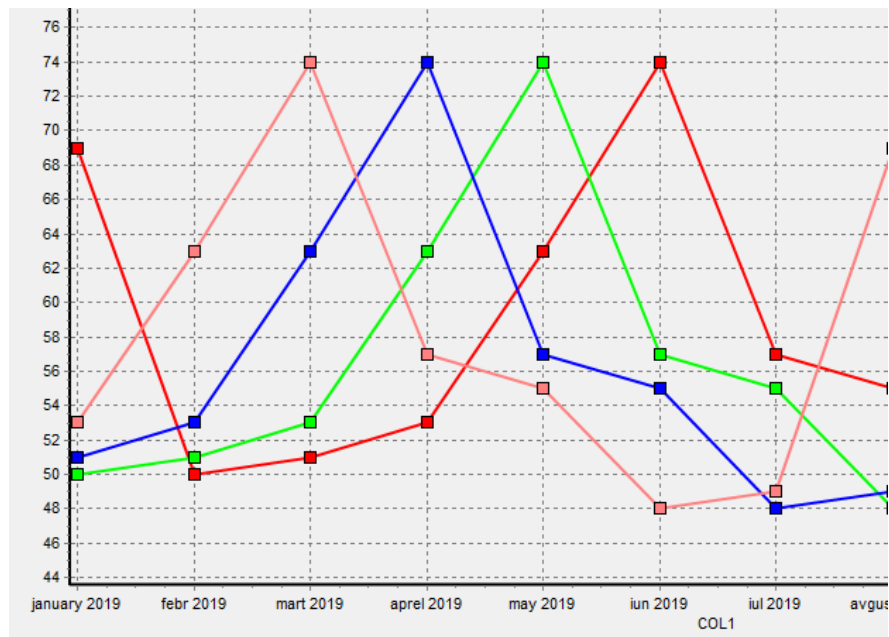


Рисунок 5.8 – Диаграмма по данным

5. После того, как нейросеть была обучена, построим прогноз с помощью обработчика “Прогнозирование”.

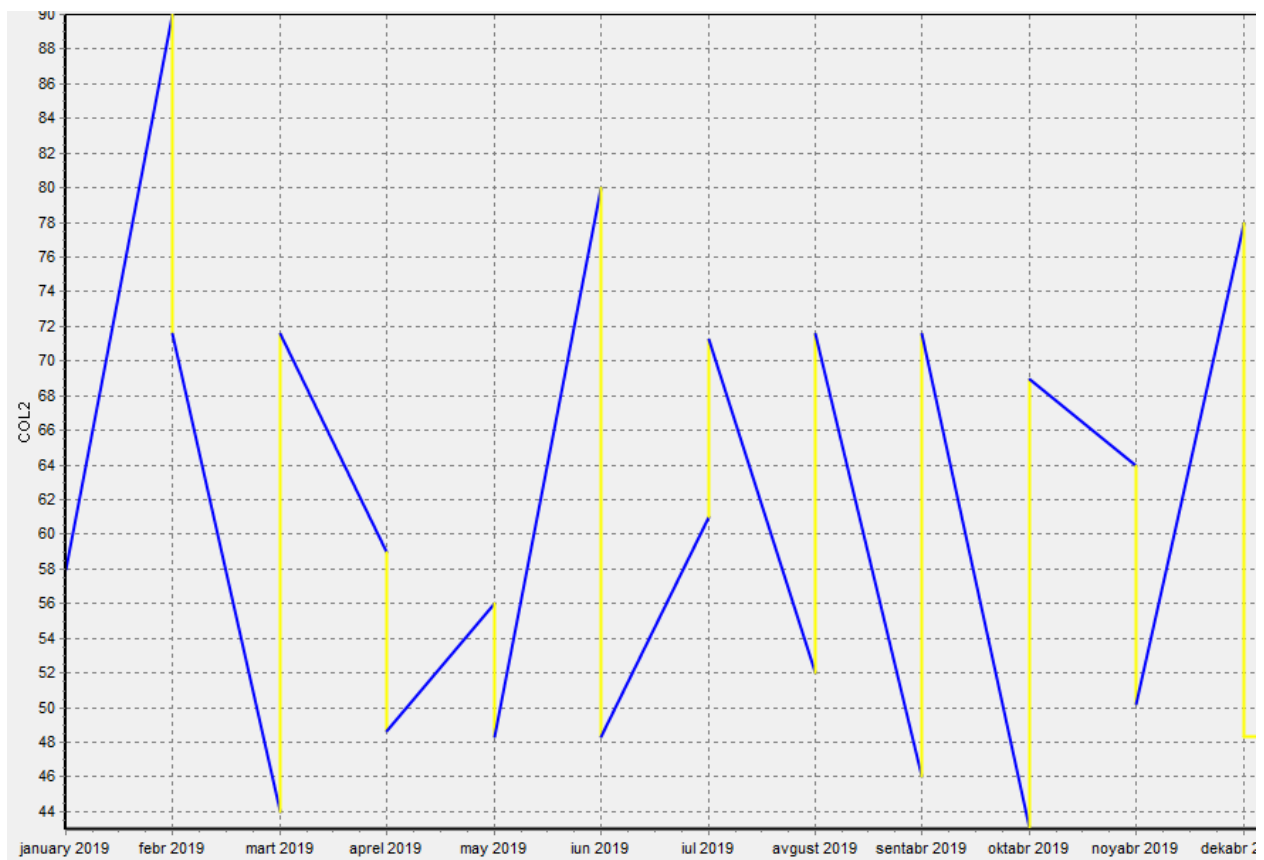


Рисунок 5.9 – Результат прогнозирования

На диаграмме прогнозирования желтой линией отображено прогнозируемое количество осадков в определенный месяц.

6 ВЫВОДЫ

В ходе выполнения расчетно-графического задания был проведен поиск ассоциативных правил для данных, представляющих собой чеки покупателей продуктового магазина. Были выявлены популярные наборы: Вода, Картофель, Колбаса, Соль, Яйца, Лук, Молоко. В визуализаторах “Правила”, “Дерево правил” и “Что-если” были определены условия и вероятности того, что приобретет посетитель магазина, если он уже купил определенную товарную позицию в магазине.

Также было проведено прогнозирование временного ряда количества осадков за 3 года. При помощи “Редактирование выбросов и экстремальных значений” и “Спектральная обработка” была проведена обработка данных от аномалий и шумов, мешающих построению дальнейшей тенденции. Для прогнозирования временного ряда при помощи нейросети было проведена обработка данных методом “Скользящее окно” с глубиной погружения 12 месяцев. Проведено обучение нейросети и построен прогноз изменения количества осадков.