МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Систем обработки информации и управления»

ОТЧЕТ

**Лабораторная работа №\_\_2\_\_**

по дисциплине«ИИ в задачах бизнес-аналитики»

Тема: «Microsoft ML Studio»

ИСПОЛНИТЕЛЬ: \_\_Журавлев Н. В.\_\_

ФИО

группа ИУ5-44М \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

"24"\_\_апреля\_\_\_2025 г.

ПРЕПОДАВАТЕЛЬ: \_\_\_\_Сухобоков А.В.\_\_\_\_\_

ФИО

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

"24"\_апреля\_\_\_2025 г.

Москва - 2025

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Постановка задачи

Шаг 1

Импортировать набор данных для регрессии в ML Studio. DATASETS => +NEW =>

Шаг 2

Создать новый эксперимент. EXPERIMENTS => +NEW => Blank Experiment

Шаг 3

Переименовать эксперимент. "Experiment created on.." => "Regression"

Шаг 4

Добавить импортированный набор данных. Перетащить ранее загруженный датасет Folds5x2\_pp.csv в область экспиримента.

Шаг 5

Визуализировать данные. Выбрать Visualize.

Шаг 6

Добавить описание к экспирименту из датасета.

Description:

Данные электростанции комбинированного типа

Шаг 7

a) Переименовать столбцы датасета. Перетащить объект Edit Metadata в область экспиримента. Search..=> Edit Metadata => Lauch columns selector

Temp,Vacuum,Pressure, Humidity,Output

б) Привести значения в датасете к типу Integer Data Type => 'Integer'

в) Запустите эксперимент, нажав RUN внизу страницы.

г) Визуализировать данные. Выбрать Visualize.

Шаг 8

В нашем наборе данных каждая строка представляет данные по электростанции комбинированного типа в конкретный момент времени за 6 лет (2006-2011 гг.), а каждый столбец - это показатели(признаки) электростанции в этот менет времени.

Выбор правильного набора признаков(фич) определяет точность прогноза. Некоторые признаки имеют сильную корреляцию с другими признакми, что приводит к переизбыточности данных, поэтому могут быть удалены.

Модуль Filter Based Feature Selection определяет признаки в наборе данных с наибольшей предсказательной силой(точностью)

a) Отфильтровать датасет, выбрав наиболее влияющие признаки на прогноз чистой почасовой выработки электроэнергии(Output).

б) Указать необходимое количество признаков. Number of desired features => '3'

в) Выбрать статистический метрику Пирсона для начисления баллов для каждого столбца(признака). Feature scoring method => 'Pearson Correlation'

г) Запустить эксперимент, нажав RUN внизу страницы.

д) Визуализировать данные. Выбрать Visualize.

Шаг 9

Для того, чтобы понять насколько алгоритм ML хорошо отрыбатывает, необходимо его протестировать. Для этого - имеющие данные необходимо разделить на выборку, на которых модель будет обучаться (обучающая выборка), и на выборку для проверки модели (тестовая выборка).

Делим датасет на обучающую и тестовую выборку в соотношении 70/30.

Шаг 10

Выбор типа модели:

а) Обучение с учителем:

линейная регрессия (выход - действительное число или числовой вектор)

Множество X – объекты, наблюдения, примеры, ситуации, входы (samples) – пространство признаков.

Множество Y – ответы, отклики, «метки», выходы (responses).

Имеется некоторая зависимость (детерминированная или вероятностная), позволяющая по x ∈ X предсказать y ∈ Y. Зависимость известна только на объектах из обучающей выборки.

Задача обучения с учителем: восстановить (аппроксимировать) зависимость, т. е. построить функцию (решающее правило) f : X → Y, по новым объектам x ∈ X приближающую y на всем множестве X.

Поскольку мы хотим предсказать показатель чистой почасовой выработки электроэнергии, который является числом, мы будем использовать алгоритм линейной регрессии.

а) Перетащить объект Linear Regression в область экспиримента.

б) Выбрать метод наименьших квадратов и его параметры.

Шаг 11

а) Обучить модель. Выбрать целевой(выходной) признак.

б) Посмотреть коэффициенты каждого из признаков.

Шаг 12

Получить предсказания модели на тестовом датасете.

a) Соединить модель с обучающимим датасетом и выбранным алгоритом. Соединить Train Model с модулем Split Data и модулем Linear Regression

б) Запустить эксперимент, нажав RUN внизу страницы.

в) Визуализировать данные.

Шаг 13

Оценить модель.

а) Перетащить объект Score Model в область экспиримента.

б) Запустить эксперимент, нажав RUN внизу страницы.

в) Визуализировать данные.

Шаг 14

Повторить шаги с 10 по 12, НО... а) Выбрать метод градиентный спуск и его параметры: размер шага, количество итераций, вес регуляризации. Solution method=> 'Online Gradien Descent'

Learning rate => '0.1'

Number of training => '1000'

L2 => '0.001'

Этот метод минимизирует ошибку на каждом шаге процесса обучения модели.

Шаг 15

а) Сравнить две получившиеся модели. Перетащить объект Evaluate Model в область экспиримента. Search..=> Evaluate Model

б) Визуализировать данные. Выбрать Visualize.

Эксперимент №2.1 - Кластеризация

Шаг 1

Создать новый эксперимент. EXPERIMENTS => +NEW => Blank Experiment

Шаг 2

Импортировать набор данных для кластризации в ML Studio. Перетащить объект Import Data в область экспиримента.

Шаг 3

Удалить строки в данных с пропущенными значениями с помощью модуля Clean Missing Data. Search..=> Clean Missing Data => Cleaning mode: 'Remove entire row'

Шаг 4

a) Переименовать столбцы датасета. Перетащить объект Edit Metadata в область экспиримента. Search..=> Edit Metadata => Lauch columns selector: All columns, All labels

F1,F2,F3,F4,Label

б) Запустите эксперимент, нажав RUN внизу страницы.

в) Визуализировать данные. Выбрать Visualize.

Шаг 5

a) Столбец Label сделать меткой. Перетащить объект Edit Metadata в область экспиримента. Search..=> Edit Metadata => Lauch columns selector: Column names: 'Label'

Fields: 'Label'

б) Запустите эксперимент, нажав RUN внизу страницы.

в) Визуализировать данные. Выбрать Visualize.

Шаг 6

Делим датасет на обущающую и тестовую выборку в соотношении 60/40. Перетащить объект Split Data в область экспиримента. Search..=> Split Data => Fraction of rows.... =>'0.7'

Шаг 7

Выбор типа модели:

а) Обучение без учителя:

кластеризация (выход - группировка данных по схожим признакам в кластеры(группы)

Есть только множество X – объекты, наблюдения, примеры, ситуации.

Задача обучения без учителя: известны только описания множества объектов (обучающей выборки), требуется обнаружить внутренние взаимосвязи, зависимости, закономерности, существующие между объектами.

Все алгоритмы кластеризации основаны на вычислении расстояния между точками в выборке.

Восьпользуемся алгоритмом k-средних. Где k - это количество центроидов. Центроид - это точка, представляющая каждый кластер, на которые мы хотим разделить наш датасет. Если ваши данные размечены, вы можете использовать значения метки, чтобы определять выбор кластеров и оптимизировать модель. Если ваши данные не имеют метки, алгоритм создает кластеры, представляющие возможные категории, основываясь исключительно на данных.

а) Перетащить объект K-Means Clustering в область экспиримента. Search..=> K-Means Clustering

Create trainer mode=> 'Single Parameter'

Number of Centroids: '2'

Metric: 'Cosine'

Iterations: '10000'

Assign Label Mode: 'Owerwrite...'

Шаг 8

а) Обучить модель. Выбрать столбцы для кластеризации.

Перетащить объект Train Clustering Model в область экспиримента и столбец Output. Search..=> Train Clustering Model => Lauch columns selector

F1,F2,F3,F4,Label

Шаг 9

Применяем алггоритм к данным и инициализируем кластеры.

a) Соединить модель с обучающимим датасетом и выбранным алгоритом. Соединить Train Clustering Model с модулем Split Data и модулем K-Means Clustering

б) Запустить эксперимент, нажав RUN внизу страницы.

в) Визуализировать данные. Выбрать Visualize.

Шаг 10

Повторить шаги с 7 по 9. Но поменяем гиперпараметры алгоритма K-Means Clustering.

Create trainer mode=> 'Single Parameter'

Number of Centroids: '3'

Metric: 'Cosine'

Iterations: '100'

Assign Label Mode: 'Owerwrite...'

Шаг 11

а) Сравнить две получившиеся модели. Перетащить объект Evaluate Model в область экспиримента. Search..=> Evaluate Modelr

б) Визуализировать данные. Выбрать Visualize.

Эксперимент №2.2 - Кластеризация \*

Повторить шаги из прошлого эксперимента (выгрузить датасет из прошлого эксперимента с помощью модуля Convert to Dataset) На шаге 7 - Выбор типа модели: применить способ подбора количества центроидов с помощью параметров алгоритма Create trainer mode=> 'Parameter Range'

User Range Builder: '2,3,4,5'

Metric: 'Cosine'

Iterations: '1000'

Assign Label Mode: 'Owerwrite...'

На шаге 8 добавить модуль Train Clustering Model(All columns Exclude column names: Class) и параллельно(для сравнения) добавить модуль Sweep Clustering (Exclude column names: Class; Metric for measuring : Simplified ...'; Specify parameter sweeping: 'Entire grid'

Сравнить два способа с помощью модуля Evaluate Model:

Модуль Sweep Clustering создает несколько моделей кластеризации, перечисленных в порядке точности. Модели измеряются с использованием всех возможных метрик.

Эксперимент №3 - Классификация

Шаг 1

Импортировать набор данных для классификациив ML Studio. DATASETS => +NEW =>

Шаг 2

Создать новый эксперимент. EXPERIMENTS => +NEW => Blank Experiment

Шаг 3

a) Выбрать необходимые столбцы датасета. Перетащить объект Select Columns in Dataset в область экспиримента. Search..=> Select Columns in Dataset => Lauch columns selector: Columns names:

STG,SCG,STR,LPR,PEG,UNS

б) Запустите эксперимент, нажав RUN внизу страницы.

в) Визуализировать данные. Выбрать Visualize.

Шаг 4

a) Столбец UNS сделать меткой и переименовать . Перетащить объект Edit Metadata в область экспиримента. Search..=> Edit Metadata => Lauch columns selector: Column names: 'UNS'

Fields: 'Label' New column names: 'Level'

б) Запустите эксперимент, нажав RUN внизу страницы.

в) Визуализировать данные. Выбрать Visualize.

Шаг 5

Выбор типа модели:

а) Обучение c учителем:

Воспользуемся алгоритмом Multiclass Neural Network - частный случай регрессии.

а) Перетащить объект Multiclass Neural Network в область экспиримента. Search..=> Multiclass Neural Network

Create trainer mode=> 'Single Parameter'

Number of hidden nodes: '1000'

The learning rate: '0.1'

Number of learning iterati: '1000'

The initial learning: '0.1'

The momentum: '0'

Шаг 6

Обучим, применим и оценим модель. Перекрестно оценем параметры модели классификации путем разделения данных с помощью модуля Cross Validate Model

a) Перетащить объект Cross Validate Model в область экспиримента. Search..=>Cross Validate Model=> Selected columns:' Column names: Level'

б) Запустите эксперимент, нажав RUN внизу страницы.

в) Визуализировать данные. Выбрать Visualize.

Шаг 7

Повторить шаги с 5 по 6. Но поменяем гиперпараметры алгоритма Multiclass Neural Network

Create trainer mode=> 'Single Parameter'

Number of hidden nodes: '100'

The learning rate: '0.1'

Number of learning iterati: '100'

The initial learning: '0.1'

The momentum: '0'

Шаг 8

а) Сравнить две получившиеся модели. Перетащить объект Evaluate Model в область экспиримента. Search..=> Evaluate Modelr

б) Визуализировать данные. Выбрать Visualize.

Эксперимент №3.2 - Класcификация\*

Применить на данных из эксперимента 3.1 алгоритмы кластеризации из эксперимента 2.2. И на полученных кластерах повторить эксперимент 3.1

# Ход работы

## Эксперимент 1

При проведении первого эксперимента загрузим файл в саму программу (рис. 1).

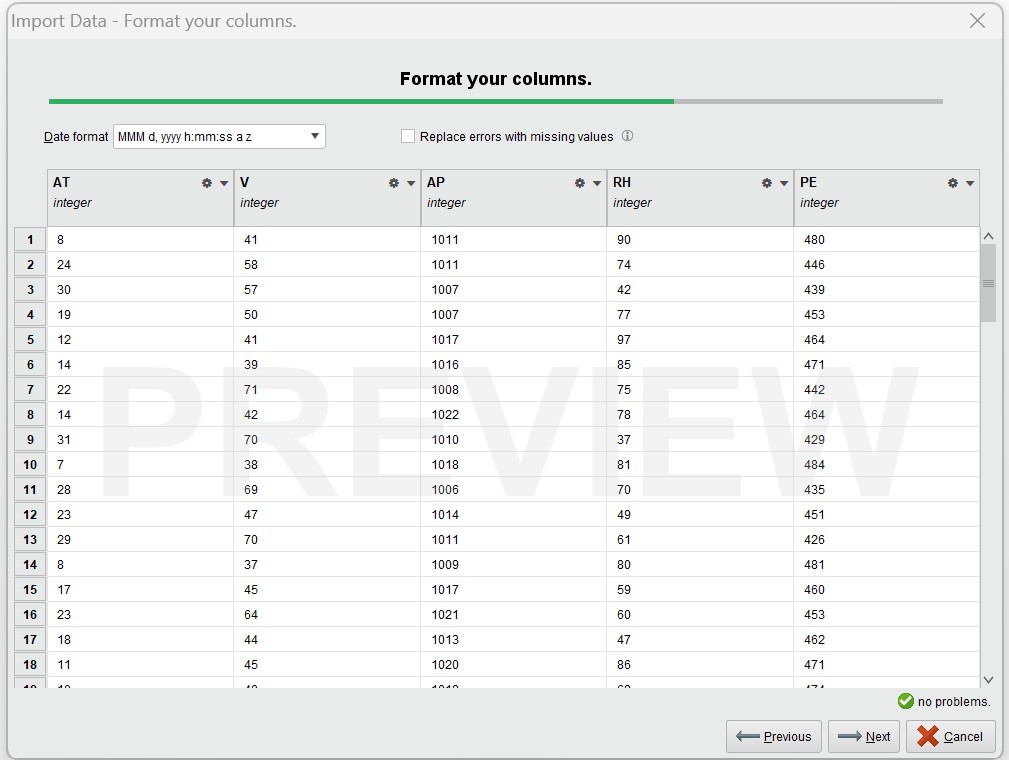


Рисунок . Набор данных

После посмотрим гистограмму распределения и диаграмму рассеивания, представленные на рис. 2 и 3.

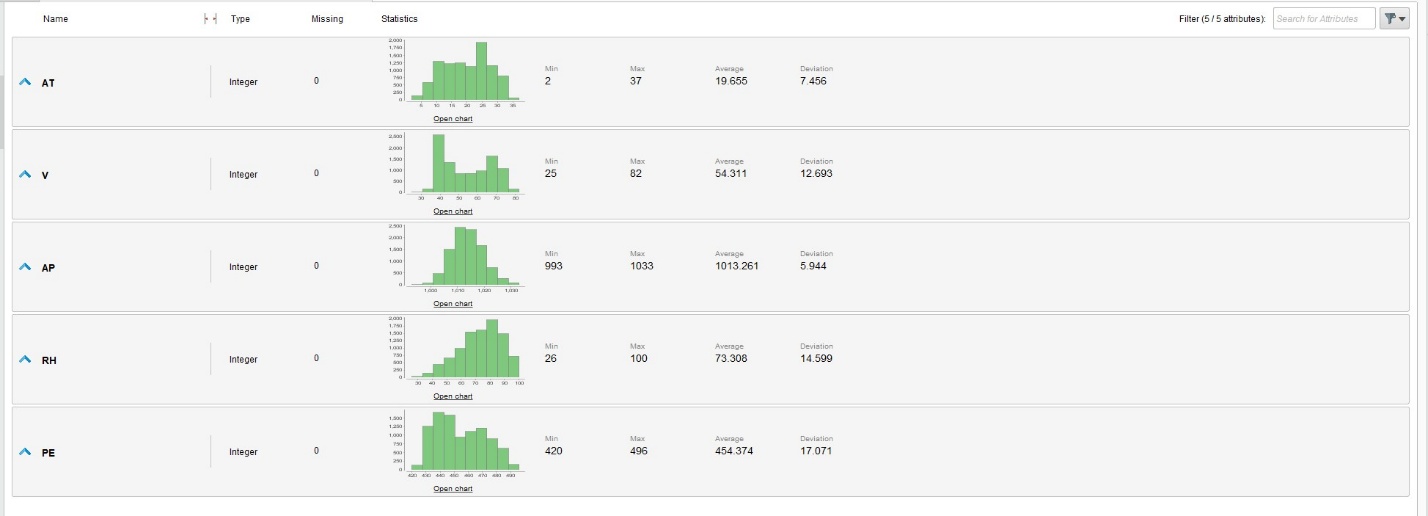


Рисунок . Гистограмма распределения

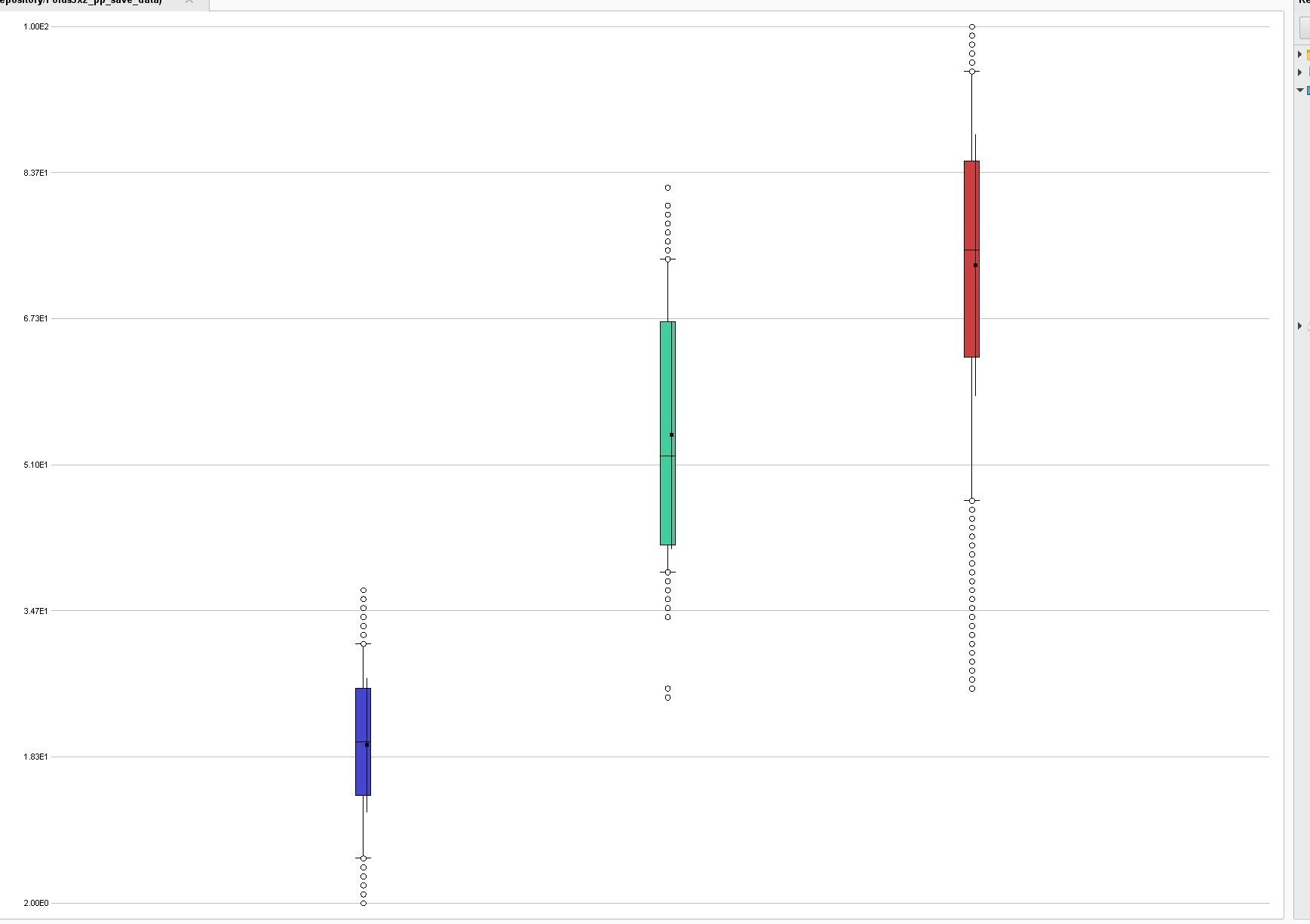


Рисунок . Диаграмма рассеивания

В рабочем пространстве поставим блок для переименования столбцов исходного набора данных (рис. 4).

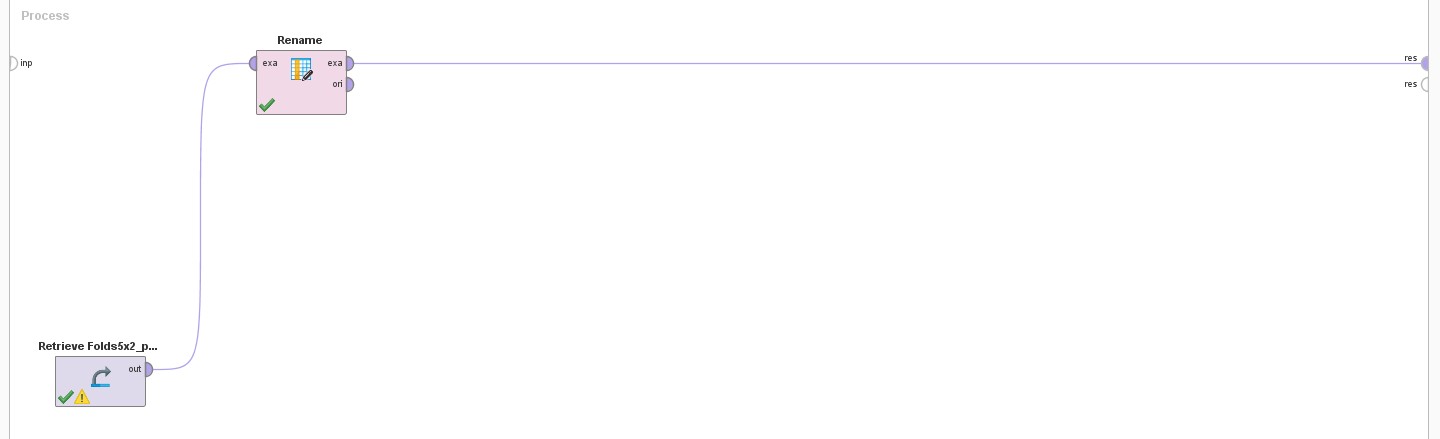


Рисунок . Добавленный блок переименования

Построим матрицу корреляции (рис. 5) и построим графики Scatter для некоторых столбцов (рис 6 - 8).

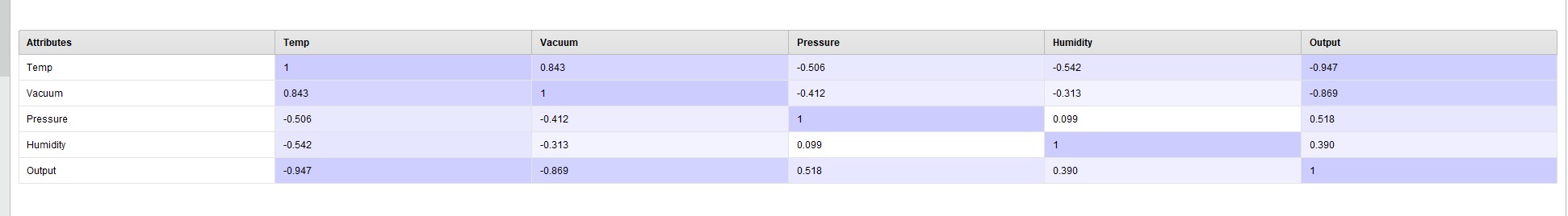


Рисунок . Матрица корреляции

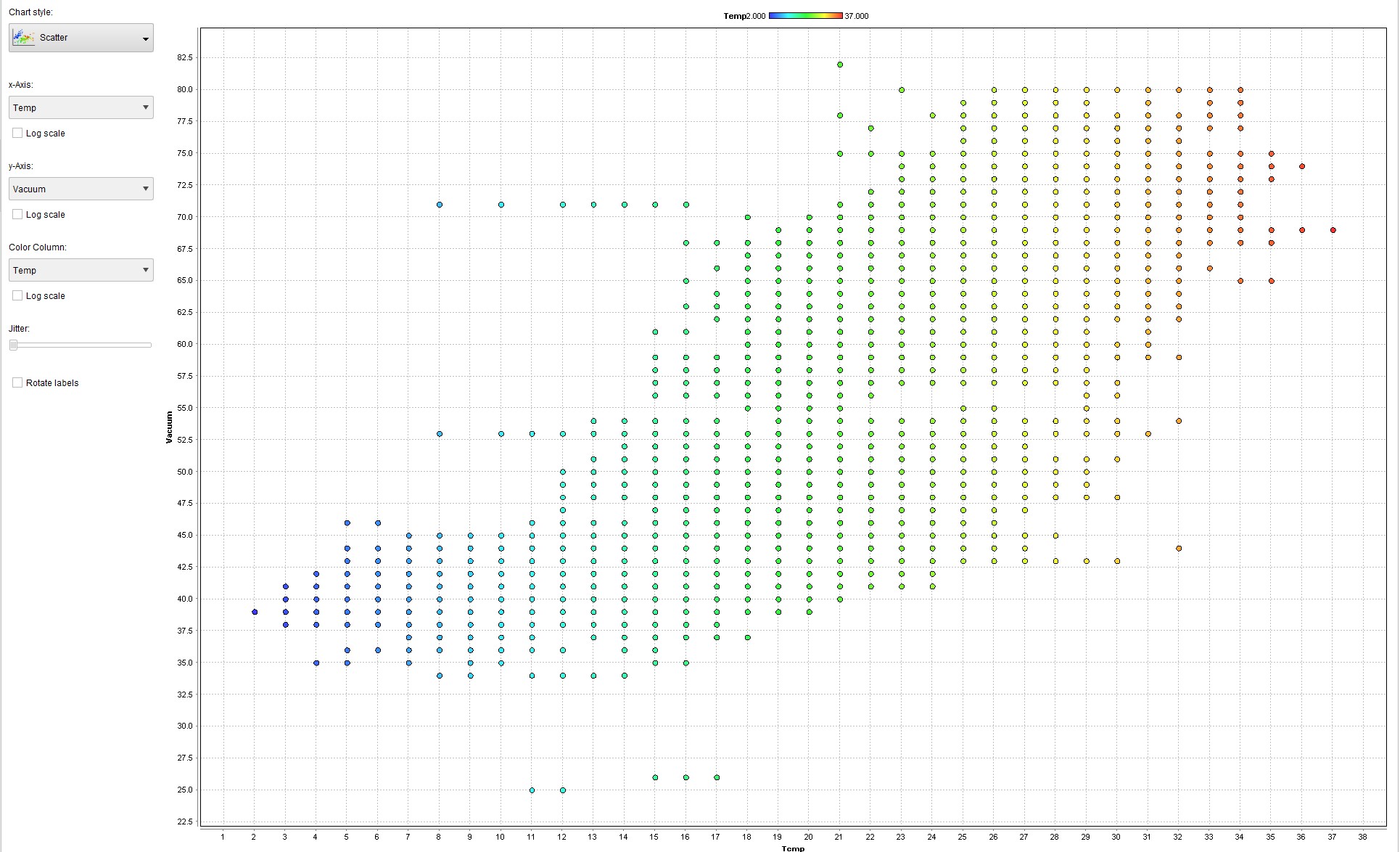


Рисунок . Связь между ЧПВ электроэнергии и Pressure

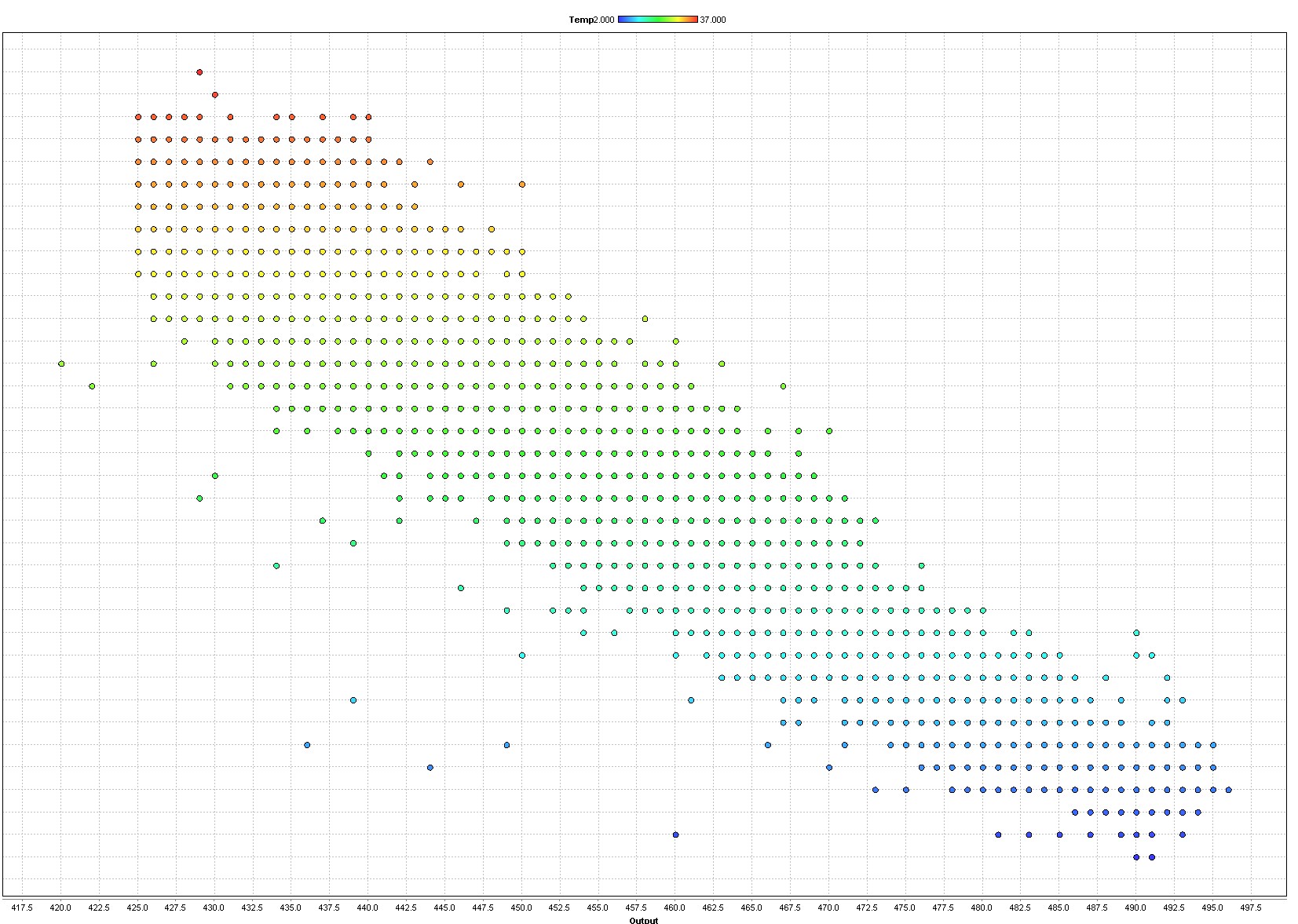


Рисунок . Связь между ЧПВ электроэнергии и Temp

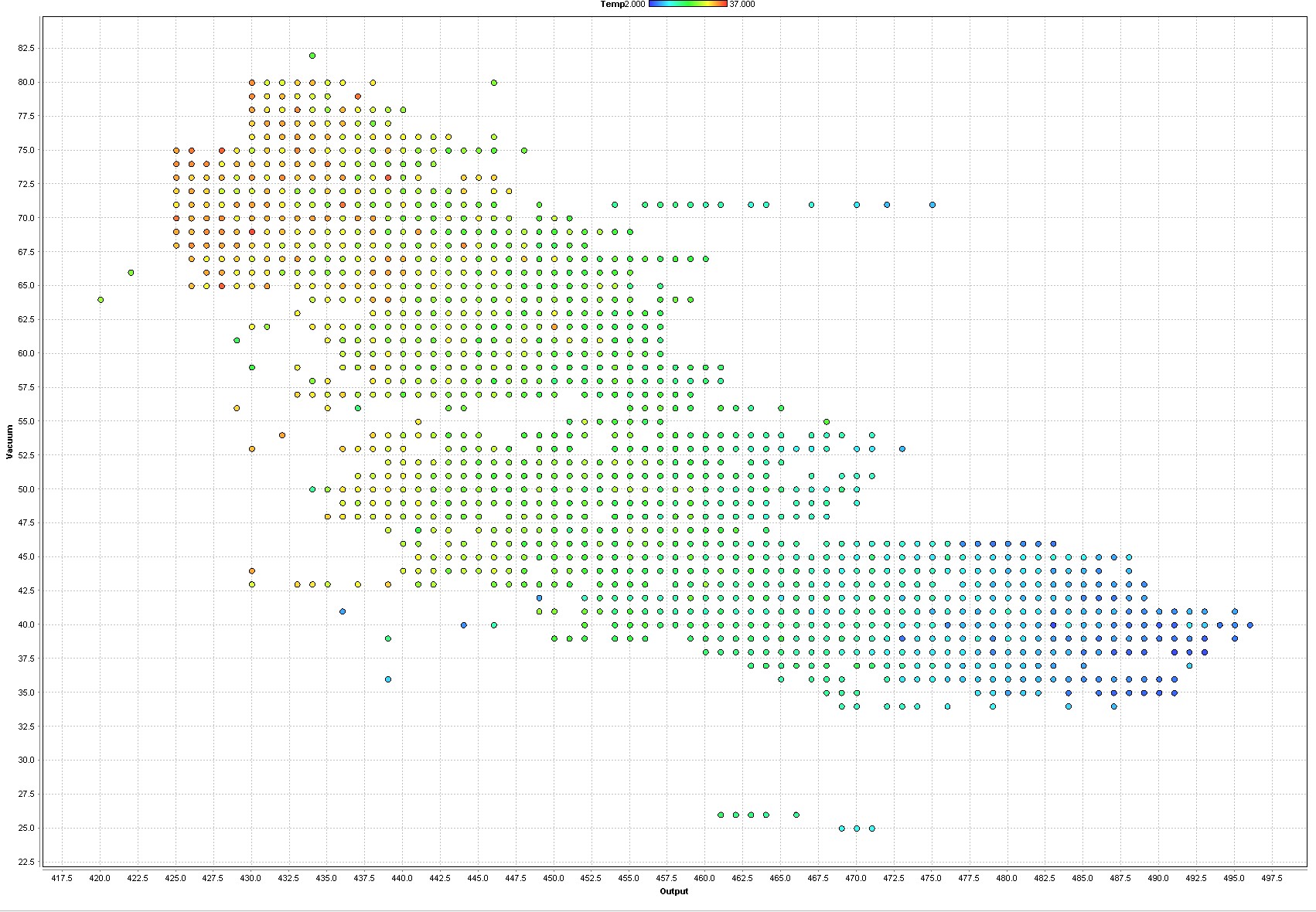


Рисунок . Связь между ЧПВ электроэнергии и Vacuum

Построим из блоков модель с линейной регрессией (рис. 9). Результат оценки модели представлен на рис. 10 и 11.

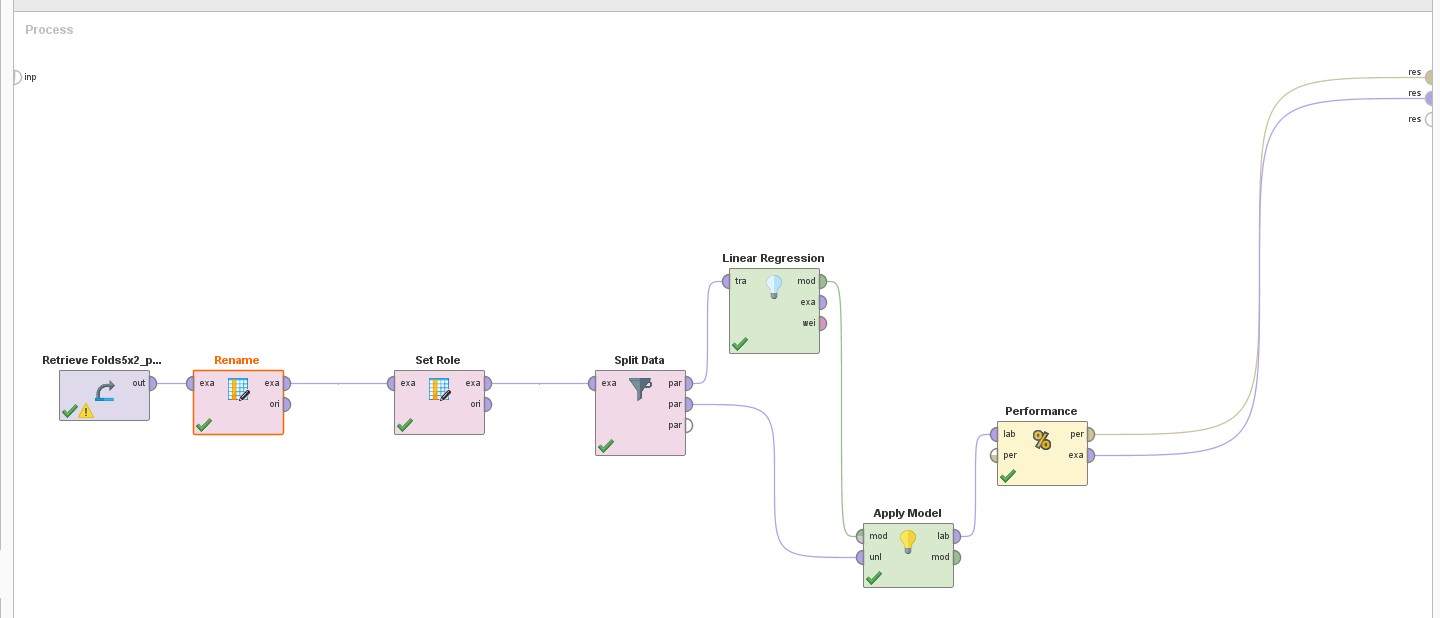


Рисунок . Модель для эксперимента 1

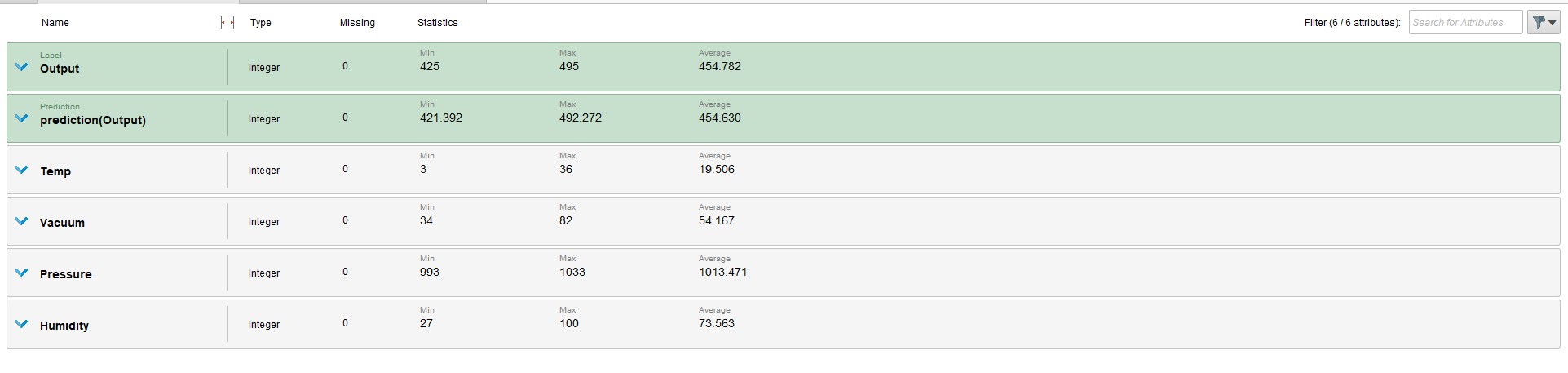


Рисунок . Результат исполнения

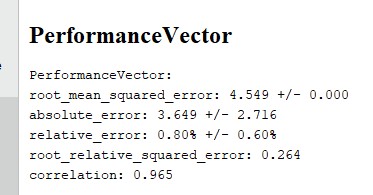


Рисунок . Результат исполнения

Построим модель с другим методом (рис. 12) и сравним их. Результаты моделей представлены на рис. 13, 14.

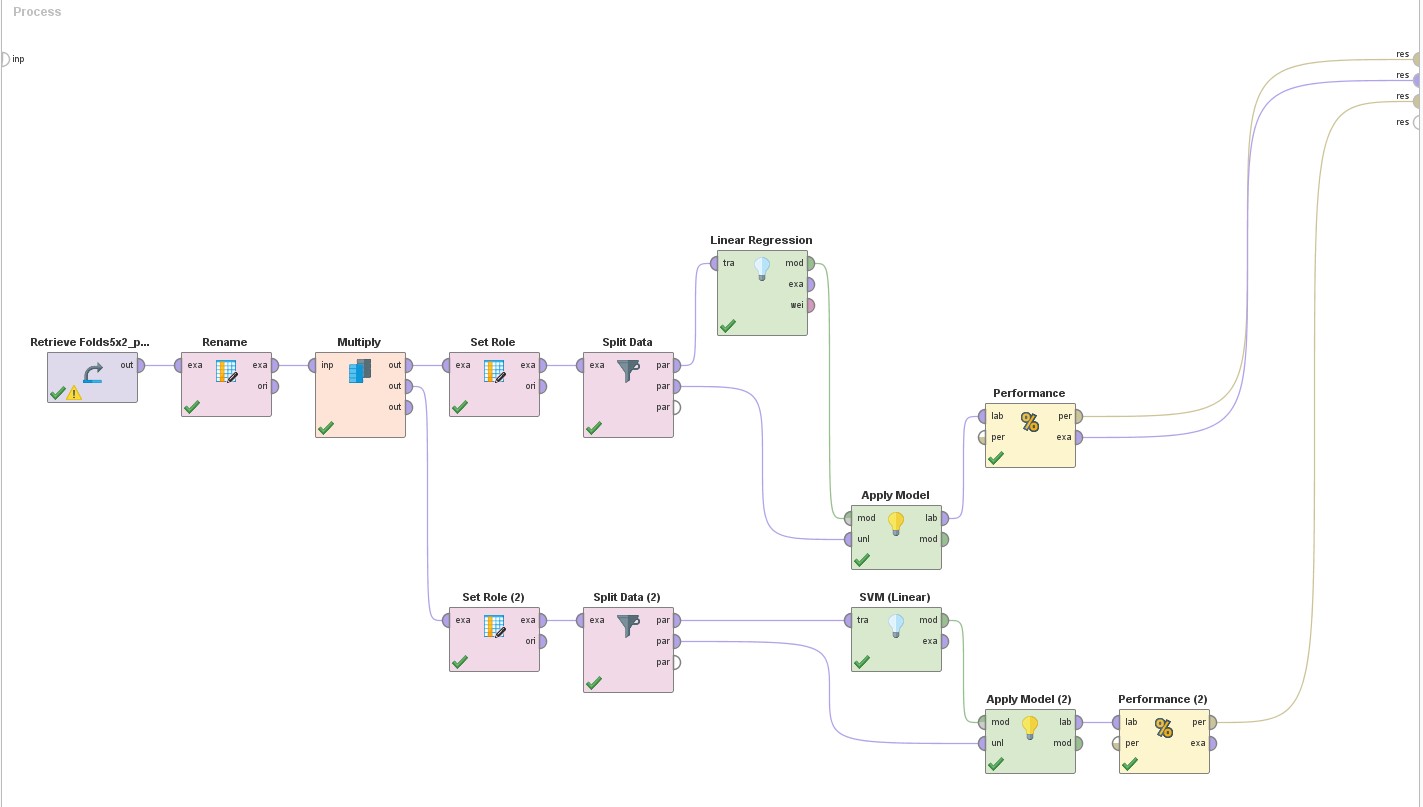


Рисунок . Две модели для первого эксперимента

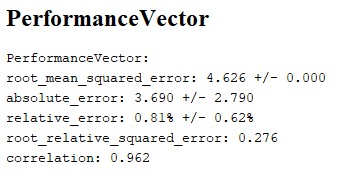


Рисунок . Результат исполнения первой модели

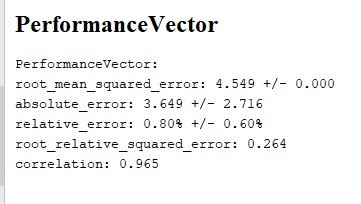


Рисунок . Результат исполнения второй модели

## Эксперимент 2.1

Для проведения необходимо построить из блоков следующую структуру, представленную на рис. 15.

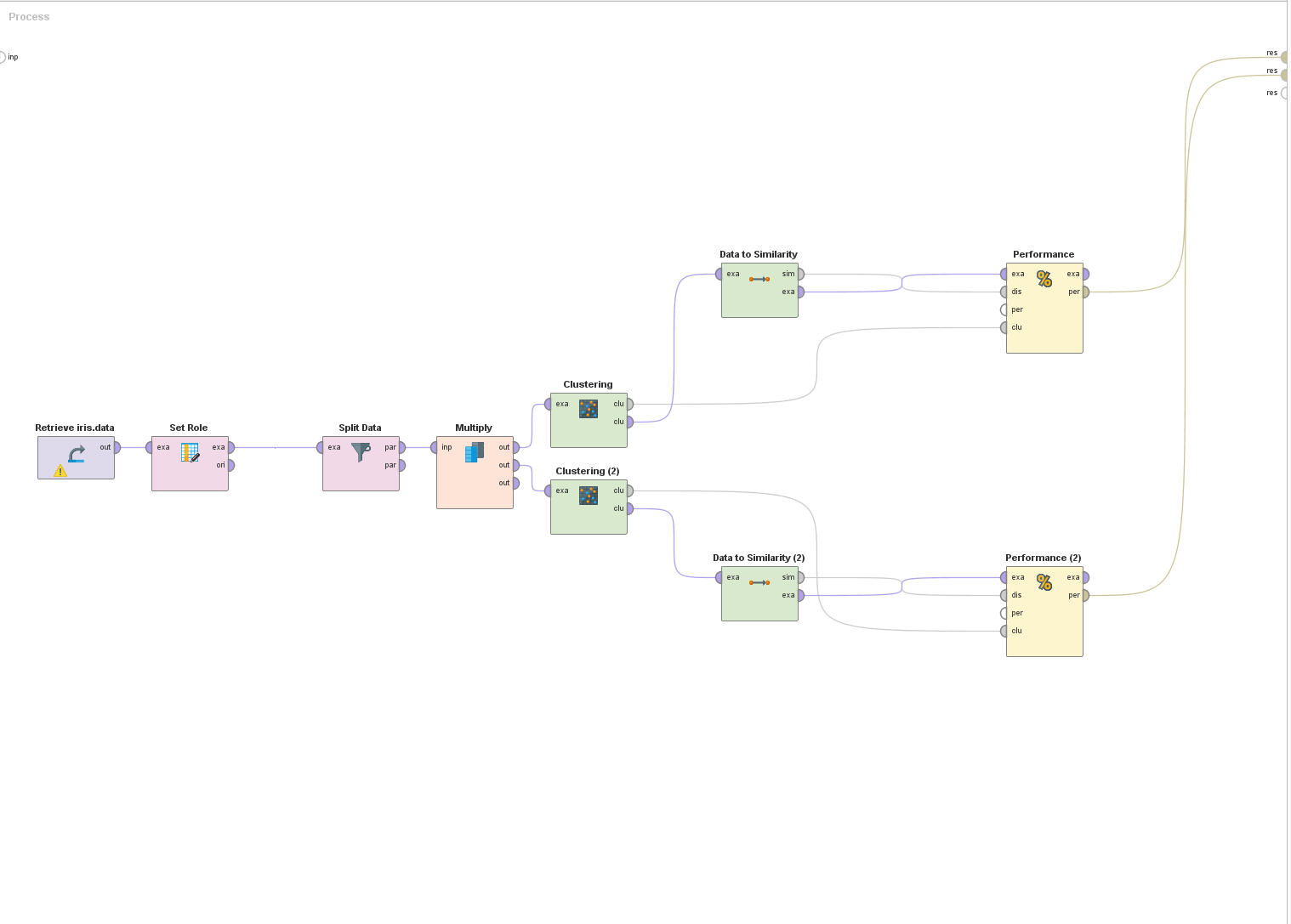


Рисунок . Модель для эксперимента 2.1

Результаты 2 моделей представлены на рис. 16-17.

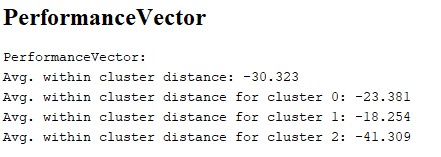


Рисунок . Результат выполнения первой модели

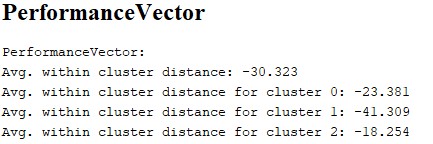


Рисунок . Результат выполнения второй модели

## Эксперимент 2.2

Для проведения необходимо построить из блоков следующую структуру, представленную на рис. 18.

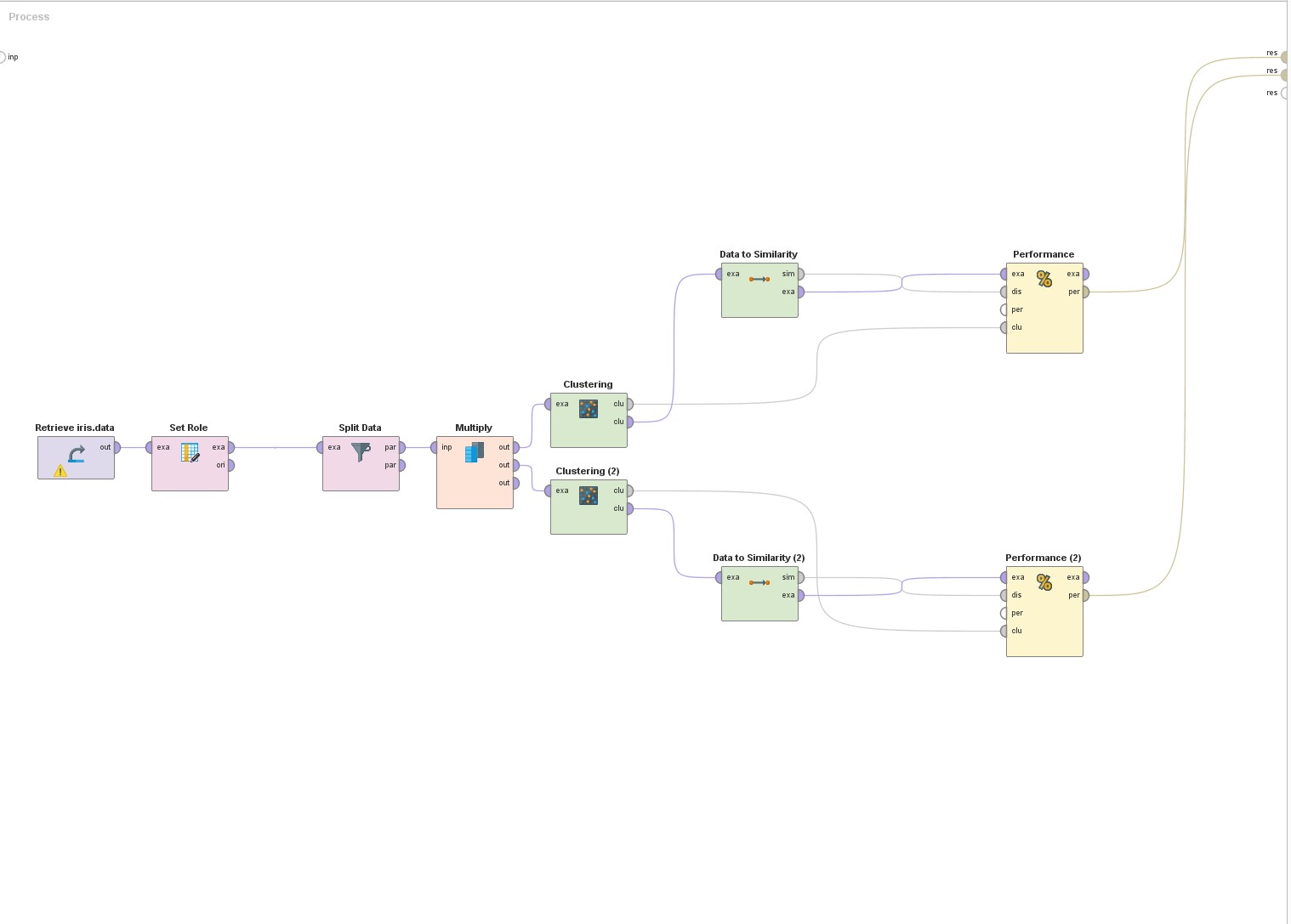


Рисунок . Модель для эксперимента 2.2

Результаты 2 моделей представлены на рис. 19-20.

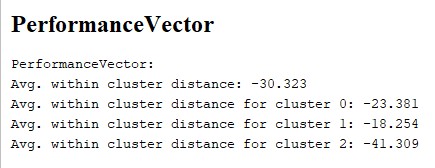


Рисунок . Результат выполнения первой модели

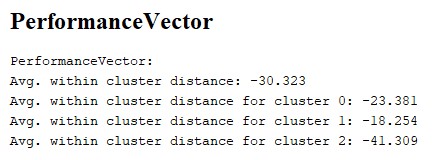


Рисунок . Результат выполнения второй модели

## Эксперимент 3.1

Для проведения необходимо построить из блоков следующую структуру, представленную на рис. 21.

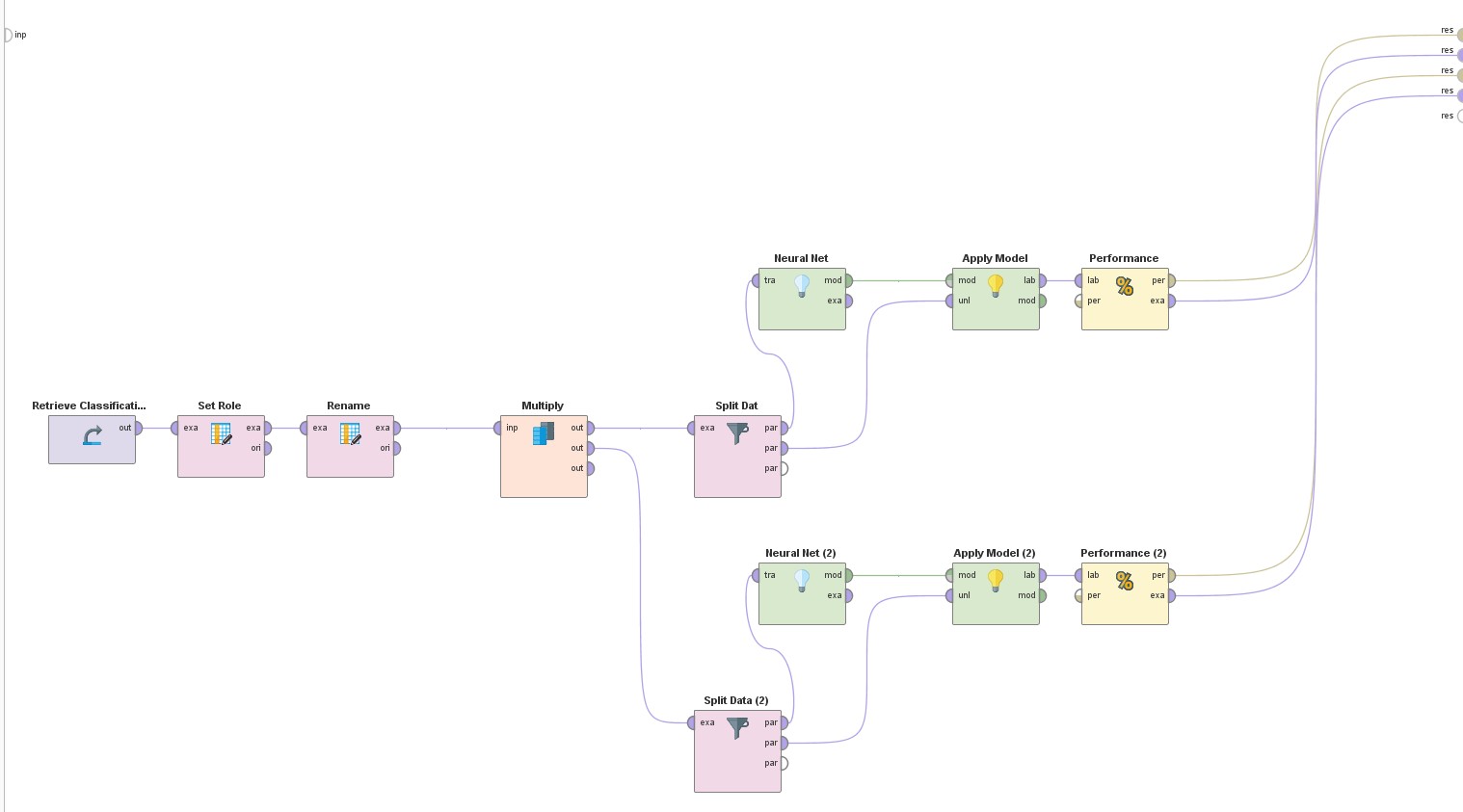
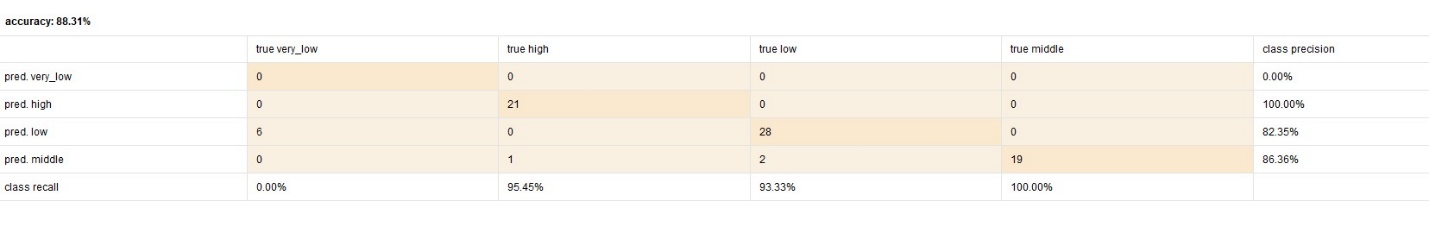


Рисунок . Модель для эксперимента 3.1

Результаты 2 моделей представлены на рис. 22-23.



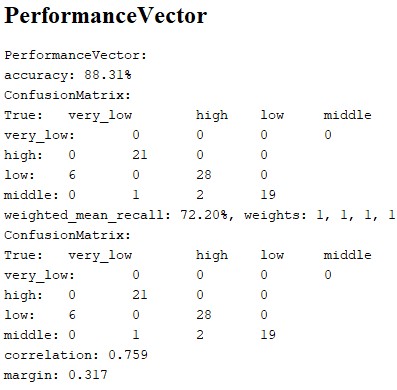
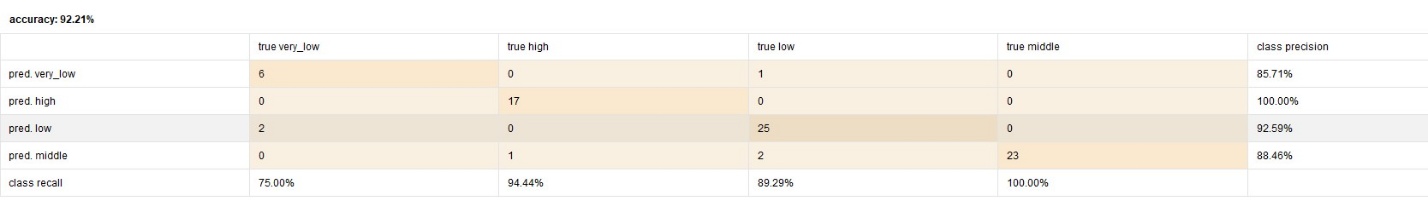


Рисунок . Результат выполнения первой модели



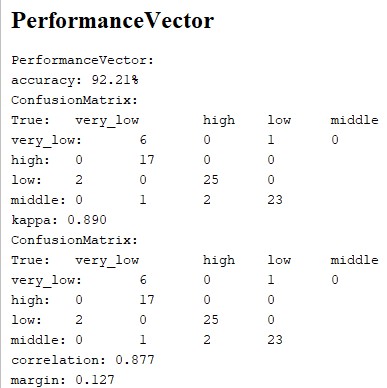


Рисунок . Результат выполнения второй модели

## Эксперимент 3.2

Для проведения необходимо построить из блоков следующую структуру, представленную на рис. 24.

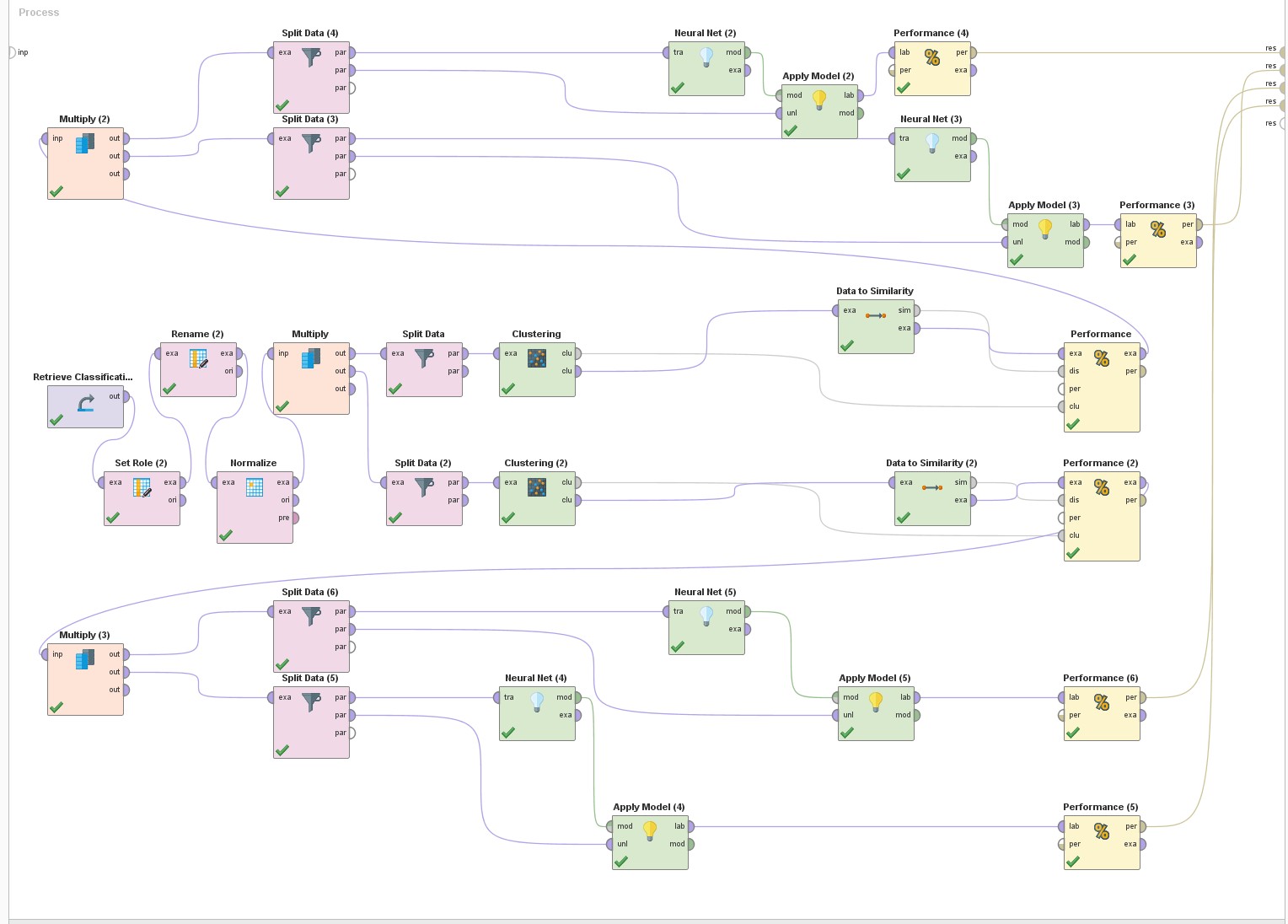


Рисунок . Модель для эксперимента 3.2

Результаты 4 моделей представлены на рис. 25-28.

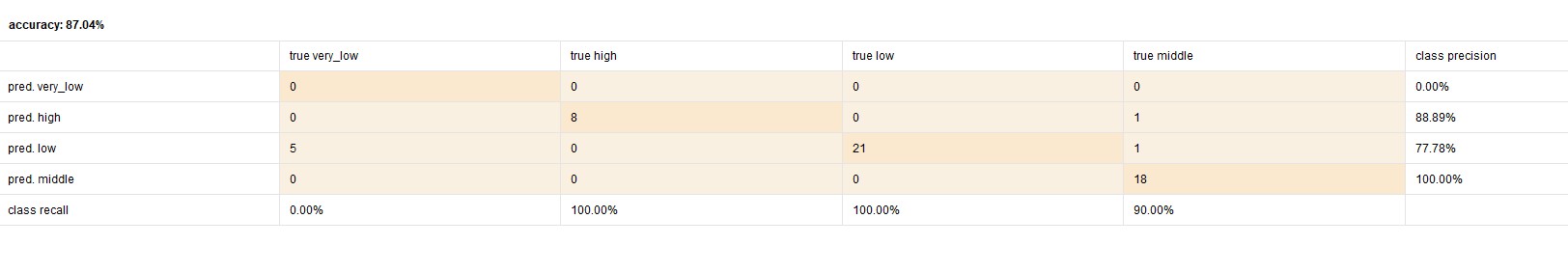
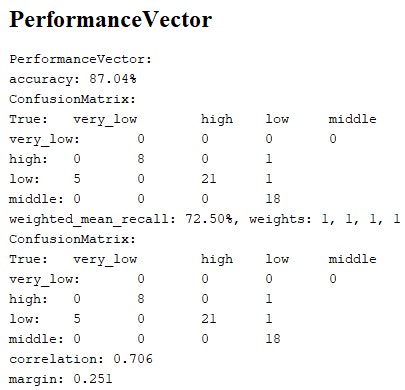


Рисунок . Результат выполнения первой модели

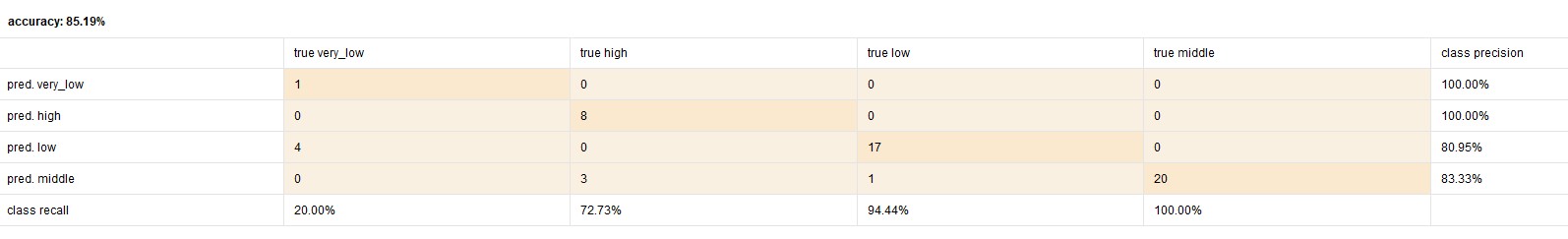
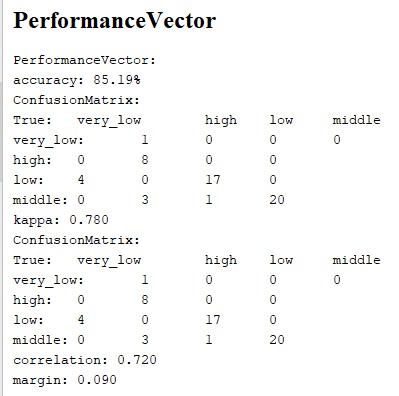


Рисунок . Результат выполнения второй модели

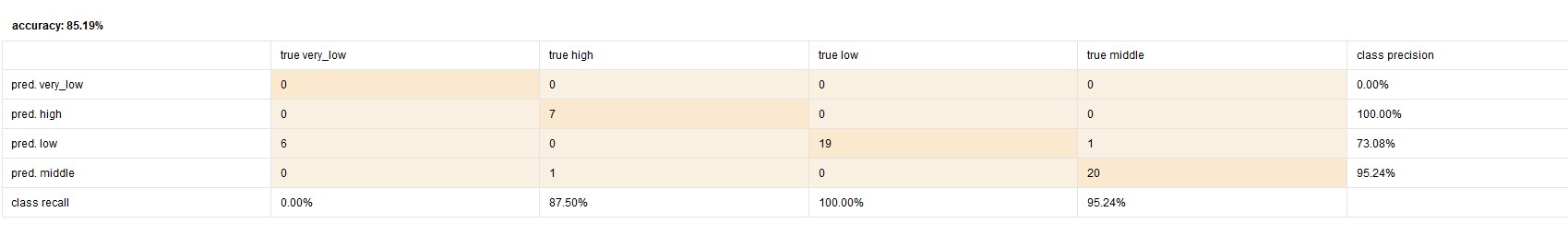
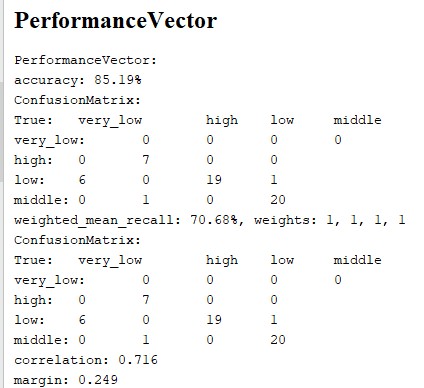


Рисунок . Результат выполнения третьей модели

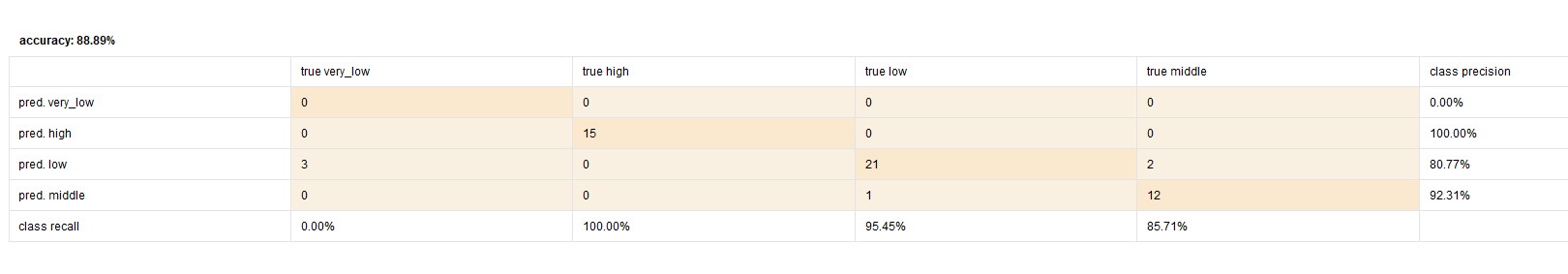
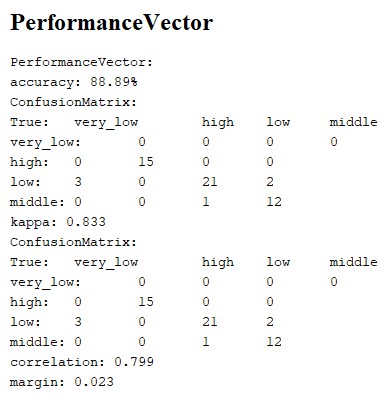


Рисунок . Результат выполнения четвёртой модели

# Вывод

В ходе выполнения лабораторной работы были получены навыки взаимодействия с Microsoft ML Studio, проведения обработки и анализа данных с помощью машинного обучения.