МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное

образовательное учреждение высшего образования

«ЧЕРЕПОВЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

|  |
| --- |
| Иформационных технологий |
| наименование института (факультета) |
| Математическое и программное обеспечение ЭВМ |
| наименование кафедры |

ОТЧЁТ

по учебной: технологической практике 1

Листов 24

Студента Микуцких Григория Андреевича группы 1ПИб-02-1оп-22

Место прохождения практики

ФГБОУ ВО «Череповецкий государственный университет» кафедра математического и программного обеспечения ЭВМ, компьютерный класс

Студент \_\_\_\_\_\_\_\_\_\_\_\_

(подпись)

Руководитель практики

от кафедры МПО ЭВМ доцент к.т.н. Юдина О.В.\_\_\_\_\_\_\_\_

(должность) (подпись, Ф.И.О.)

\_2023\_ год

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 3](#_Toc153299094)

[1. KNIME - «Konstanz Information Miner» 3](#_Toc153299095)

[2. SPSS 4](#_Toc153299096)

[3. SAS 5](#_Toc153299097)

[4. STATISTICA 6.0 6](#_Toc153299098)

[5. Minitab 8](#_Toc153299099)

[6. R + R-Studio 9](#_Toc153299100)

[7. Python 11](#_Toc153299101)

[8. Orange 12](#_Toc153299102)

[9. Loginom 14](#_Toc153299103)

[РАЗДЕЛ 1: РАБОТА В ПАКЕТЕ ORANGE, ЕГО ВОЗМОЖНОСТИ, РЕЗУЛЬТАТЫ ВЫПОЛНЕНИЯ ЗАДАНИЯ 16](#_Toc153299104)

[РАЗДЕЛ 2: ОПИСАНИЕ РЕШЕНИЯ ОПТИМИЗАЦИОННОЙ ЗАДАЧИ В MICROSOFT EXCEL 22](#_Toc153299105)

[ЗАКЛЮЧЕНИЕ 23](#_Toc153299106)

[СПИСОК ЛИТЕРАТУРЫ 24](#_Toc153299107)

# ВВЕДЕНИЕ

Навыки работы с данными необходимы при работе в таких областях как Data Mining, Big Data, машинное обучение. Для решения данных задач в настоящее время создано много пакетов прикладных программ, позволяющих получать оценки, делать выводы, проводить необходимые расчеты.

Цель технологической практики: обзорное знакомство с некоторыми из этих пакетов.

## KNIME - «Konstanz Information Miner»

Knime Analytics Platform – open source (то есть имеет открытый исходный код) фреймворк для анализа данных. Позволяет реализовывать полный цикл анализа данных, включающий чтение данных из различных источников, преобразование и фильтрацию, анализ, визуализацию и экспорт.

Основные возможности KNIME:

* ‌читать данные из различных источников;
* ‌объединять и пересекать эти данные;
* ‌строить различные цепочки обработки этих данных;
* ‌создавать интерактивные визуализации данных;
* ‌использовать машинное обучение;
* ‌использовать готовые решения сообщества KNIME для работы с данными.

KNIME поддерживает различные форматы данных, включая текстовые файлы, базы данных, файлы CSV, таблицы Excel, изображения и другие. Это позволяет пользователям работать с данными из разных источников и форматов, объединяя их в рабочих процессах.

KNIME позволяет визуализировать данные с помощью различных графических инструментов. Это позволяет пользователям лучше понять структуру данных, выявить закономерности и тренды, а также идентифицировать выбросы и аномалии. KNIME предоставляет узлы для создания графиков, диаграмм, деревьев решений и других визуальных элементов.

Он хорошо подойдет для людей не сильно знакомых с программированием (хотя иногда написание кода необходимо, но минимально, если вы хотите добавить большей сложности в свой рабочий процесс). С его помощью можно быстро создавать простые и средней сложности workflow – визуализированные потоки задач – и предоставлять к ним доступ через REST – архитектурный стиль взаимодействия компонентов распределённого приложения в сети. Вместо написания кода пользователь может создавать рабочие процессы, перетаскивая и соединяя готовые компоненты.

Процесс программирования логики осуществляется через создание Workflow. Workflow состоит из узлов которые выполняют ту или иную функцию (например чтение данных из баз данных, трансформация, визуализация). Узлы, соответственно, соединяются между собой стрелочками, которые показывают направление движение данных.

Workflow состоит из узлов (или «нод»). Практически у каждого узла есть конфигурационный диалог, в котором можно настраивать свойства.

Поддерживается следующие типы узлов: IO — ввод/вывод данных (например чтение CSV), Manipulation – преобразование данных (включая фильтрацию строк, столбцов, сортировку), Views – визуализация данных (построение различных графиков включая Histogram, Pie Chart, Scatter Plot, etc), Database – возможность подключения к базе данных, чтения/записи, Workflow Control – создание циклов, итерирование групп в ходе выполнения workflow и прочее.

Из узлов реализующих анализ данных доступны различные статистические методы (включая линейную корреляцию, проверку гипотез, классификацию, регрессию, кластеризацию, ассоциативные правила и другие), а также Data Mining методы (например, нейронные сети, построение decision trees, cluster view).

Фреймворк хорош для работы со студентами, поскольку наглядно видно все что происходит с данными, по каким веткам они перемещаются и как преобразуются. Студенты могут изучать реализацию существующих узлов, дописывать свои компоненты (узлы) и пополнять ими библиотеку.

Кроме того, как указывалось ранее, KNIME является приложением с открытым исходным кодом, а это означает, что оно бесплатное [1][2][3].

## SPSS

SPSS Statistics («Statistical Package for the Social Sciences» – «статистический пакет для общественных наук») — это компьютерная программа для статистической обработки данных, один из лидеров рынка в области коммерческих статистических продуктов, предназначенных для проведения прикладных исследований в общественных науках.

Она широко используется в научных исследованиях, социологии, маркетинге, бизнес-аналитике и других областях, где требуется проведение статистического анализа данных.

Пакет предназначен в первую очередь для статистиков-профессионалов. Он включает развитый аппарат статистического анализа, соизмеримый по мощности с SAS.

SPSS имеет удобные графические средства (более 50 типов диаграмм), а также развитые средства подготовки отчетов. Аналитические параметры отображаются на экране в виде простых и понятных меню и диалоговых окон. Новая контекстно-ориентированная справочная система содержит пошаговые инструкции для наиболее важных операций.

Основные функциональные возможности:

* ‌описательная статистика – это вычисляемые показатели, которые отражают значение переменных;
* ‌однофакторный дисперсионный анализ данных – статистическая значимость различий между значениями в нескольких выборках;
* ‌кластерный анализ данных;
* ‌дискриминантный анализ данных.

SPSS Statistics предоставляет пользователю мощные инструменты для обработки и анализа данных. Она позволяет импортировать данные из различных источников, таких как базы данных, электронные таблицы и текстовые файлы. После импорта данных, пользователь может проводить различные статистические анализы, включая описательную статистику, корреляционный анализ, регрессионный анализ, анализ дисперсии и многие другие.

Тарифы: 1200 долларов в год или 99 в месяц. Есть версия для студентов и преподавателей со скидкой. Есть бесплатная пробная версия на 30 дней [4].

## SAS

Пакет SAS (Statistical Analysis System) – профессиональный статистический пакет, разработанный компанией SAS Institute Inc. Последняя версия позиционируется как статистический пакет для аналитической работы на современных предприятиях.

Разработчики утверждают, что пакет эффективно работает с данными различных типов:

* маркетинговые базы данных;
* результаты клинических испытаний;
* медико-санитарных обследований;
* исследований предпочтений потребителей;
* исследований рынка ценных бумаг и прочие.

Встроенные инструменты статистического анализа могут быть применены для решения широкого круга вопросов, относящихся к различным областям деятельности, например:

* ‌промышленное производство;
* ‌телекоммуникации (проведение исследования рынка для выявления предпочтений потребителей и моделей выбора новых услуг);
* ‌правительство (применение статистических выборочных методов для разработки опросов общественного мнения);
* ‌исследование окружающей среды (построение моделей загрязнения воздуха с использованием пространственной статистики);
* ‌биотехнологии;
* ‌розничная торговля (моделирование поведения потребителей для выявления потенциальных целевых рынков для новых проектов в области электронной коммерции).

Пакет предоставляет пользователям следующие возможности статистического анализа данных:

* ‌смешанные модели;
* ‌анализ категорийных данных;
* ‌байесовский анализ;
* ‌многомерный анализ;
* ‌анализ выживаемости;
* ‌психометрический анализ;
* ‌анализ пропущенных значений.

Пакет позволяет проводить следующую работу с данными:

* ‌ввод, восстановление, модификация и управление данными;
* ‌написание отчетов и подготовка графиков;
* ‌статистический анализ;
* ‌бизнес планирование, прогнозирование и поддержка принятия решений;
* ‌исследование операций и управление проектами;
* ‌совершенствование качества;
* ‌разработка приложений.

Подробнее о возможностях SAS можно узнать на их официальном сайте: <https://www.sas.com/en_us/software/stat/features-list.html>

Достоинства SAS:

* ‌быстрая обработка очень больших объемов данных;
* ‌имеется возможность преобразования математических формул в программный код;
* ‌создание пользовательских модулей;
* ‌получение консультативной помощи в выборе методов анализа и в интерпретации его результатов, а также рекомендаций по дальнейшей работе с исходными данными;
* ‌совместимость с операционными системами Linux и Windows.

Недостатки SAS:

* ‌сложен в освоении для неспециалистов;
* ‌высокая стоимость‌.

Есть 14-дневный пробный период. Цена: 8000 долларов в год [5].

## STATISTICA 6.0

Система Statistica – это универсальная интегрированная система, предназначенная для статистического анализа и обработки данных.

Язык программирования системы Statistica 6.0 – Visual Basic – позволяет создавать дополнительные приложения, реализующие как простые преобразования данных, так и сложные вычислительные процедуры.

Пакет предоставляет инструменты для статистического и графического анализа, прогнозирования, data mining, создания собственных пользовательских приложений, интеграции, совместной работы, web-доступа и др.

STATISTICA включает графический модуль, содержащий более десяти тысяч типов графиков, имеющих разнообразные возможности редактирования, интерактивной настройки (вращение, масштабирование, прозрачность и др).

Система позволяет напрямую производить импорт/экспорт данных из Microsoft Office, работать в Microsoft Excel "внутри" STATISTICA, автоматически сохранять результаты в Microsoft Word.

Система взаимодействует с любыми реляционными базами данных (Oracle, MS SQL Server, Informix, Access и др.), хранилищами бизнес-информации SAP Business Warehouse. Имеет интеграцию с языком R, взаимодействует с Microsoft SharePoint, позволяет выгружать данные из OSI PI, поддерживает OLAP технологии.

Корпоративные версии STATISTICA полностью web-интегрированы: "ввод" и "вывод" данных через Web, направление результатов на Web-сервер, построение сложных автоматизированных систем, работающих с данными из внешних источников, проведение анализов и обновление содержания HTML-страниц на Web-сервере.

Базовые модули позволяют осуществить первичный анализ данных и подготовить аналитический отчет. Раздел включает все графические инструменты системы Statistica 6.0, а также следующее:

* ‌описательные статистики;
* ‌парные и частные корреляции;
* ‌группировку и однофакторный дисперсионный анализ;
* ‌таблицы частот, сопряженности, флагов и заголовков, многомерных откликов;
* ‌множественную регрессию;
* ‌дисперсионный анализ;
* ‌непараметрические статистики и др.

Модуль "Углубленные методы анализа" позволяет строить сложные объяснительные модели зависимостей и прогноза с помощью классических моделей временных рядов. Все анализы снабжены разнообразными интерактивными графиками и встроенным редактором Visual Basic. Данный модуль включает:

* ‌общие линейные модели;
* ‌обобщенные линейные и нелинейные модели;
* ‌общие регрессионные модели;
* ‌общие модели частных наименьших квадратов;
* ‌компоненты дисперсии;
* ‌множественную нелинейную регрессию;
* ‌временные ряды и прогнозирование;
* ‌моделирование структурными уравнениями и др.

В состав модуля "Многомерный разведочный анализ" входят:

* ‌кластерный анализ;
* ‌факторный анализ;
* ‌анализ главных компонент и классификация;
* ‌канонический анализ;
* ‌надежность и позиционный анализ;
* ‌деревья классификации;
* ‌анализ соответствий;
* ‌многомерное шкалирование;
* ‌дискриминантный анализ;
* ‌общие модели дискриминантного анализа [6].

## Minitab

Флагманский продукт компании, Minitab Statistical Software, представляет собой статистический пакет для решения самых разнообразных задач анализа данных. Он предлагает своим пользователям простой интерфейс в сочетании с многочисленными интерактивными подсказками, которые упрощают работу, помогают в интерпретации результатов и ускоряют освоение программы.

Minitab содержит все необходимые инструменты для первичной обработки данных, анализа результатов и подготовки презентационных материалов и отчетов. Пользователи легко могут корректировать входные данные, обрабатывая пропуски или опечатки, удалять лишние пробелы и контролировать размер данных, импортируемых из MS Excel, баз данных или других источников. При изменении данных все графики автоматически обновляются, так что пользователю не нужно беспокоиться об их актуальности.

Возможности по форматированию данных внутри рабочих документов позволяют легко выделять подмножества часто встречающихся значений, выбросов, бракованных изделий и многого другого. Все графические материалы, созданные в Minitab, легко можно экспортировать в MS Word или MS PowerPoint для создания презентаций и отчетов.

Пакет может выполнять большой объём статистического анализа, включая:

* базовую описательную и непараметрическую статистику;
* корреляцию;
* регрессию и логистическую регрессию;
* многомерный анализ и т.д.

Возможности построения графиков также обширны. Minitab может создавать графики временных рядов, запускать контрольные карты, точечные и ящичные диаграммы.

Кроме того, он позволяет пользователям создавать случайные наборы данных, соответствующие определенным параметрам распределения. Minitab часто используется для проверки данных на нормальность, что является необходимым условием для многих статистических инструментов.

Пакет развивается более 20 лет и широко известен в США, где он является одним из основных учебных пакетов.

MINITAB хорошо продуман по разделу описательной (дескриптивной) статистики, хорошо сконструирован и управляется с помощью удобного меню, или, по желанию пользователя, через команды, составлять которые помогают диалоговые окна пакета.

Часто используемые команды можно запускать по их первой букве. Общее число команд превышает 200. Можно составлять специальные макросы для выполнения последовательностей команд.

Импорт/экспорт данных из других Windows-приложений делается через стандартный буфер обмена. В пакете имеются разнообразные возможности по управлению данными. Пользователь при исследовании качества жизни может легко и быстро решать практически все типовые задачи, в основном из области получения описательных статистик и сравнения групповых средних, анализа временных рядов. Если на этапе создания и валидации опросника исследования качества жизни требуется применение методов многомерной статистики, то Minitab позволяет находить главные компоненты, проводить стандартный линейный или даже квадратичный дискриминантный анализ, использовать алгоритмы факторного и кластерного анализа.

Кроме того, пакет позволяет получать множество полноцветных графиков. В плане характеристики мощность Minitab достаточно силен и разнообразен, поэтому говорят, что первые четыре буквы пакета скорее надо поменять на «Махі».

Цены: для одного пользователя – 1595 долларов США (новая лицензия) или 895 долларов США (обновление существующей лицензии) [7][8][9].

## R + R-Studio

Язык программирования R является инструментом для статистической обработки данных и работы с графикой. Это программная среда с открытым исходным кодом, развиваемая в рамках GNU-проекта. R применяется там, где нужна работа с большими данными.

На данный момент насчитываются десятки качественных статистических пакетов, среди которых явными лидерами являются SPSS, SAS и MatLab. Однако, в 2013 году, несмотря на высокую конкуренцию, R стал самым используемым программным продуктом для статистического анализа в научных публикациях. Кроме того, в последнее десятилетие R становится все более востребованным и в бизнес-секторе: такие компании-гиганты, как Google, Facebook, Ford и New York Times активно используют его для сбора, анализа и визуализации данных.

В силу этого он имеет свои плюсы и минусы. Стоит отметить, что работа с языком R требует освоения некоторых навыков программирования и для использования возможностей надо потратить отдельное время на освоение. Стоит отметить, что R – это один из путей освоения и использования всевозможных методов обработки данных – от самых базовых до самых сложных и новых. Освоения данного языка позволяет работать с данными, получаемыми с помощью сложных методов, таких как ЭЭГ, МРТ, эффективно обрабатывать большие данные, использовать разнообразные методы структурного моделирования, машинного обучения и других современных подходов к анализу данных. Можно как самостоятельно реализовать практически любую статистическую процедуру, так и воспользоваться готовыми решениями, созданными сообществом и распространяемыми в форме пакетов (packages).

Работа в R заключается в написании и выполнении набора команд, которые позволяют проводить различные операции и процедуры статистической обработки данных. Выполнить команду можно двумя способами: ввести её в консоль и нажать Enter или набрать в окне скрипта и затем для выполнения нажать сочетание клавиш ctrl-Enter или кнопку Run в верхнем правом углу этой области.

Достоинства среды R:

* бесплатная и кроссплатформенная;
* богатый арсенал стат. методов;
* качественная векторная графика;
* более 7000 проверенных пакетов;
* гибкая в использовании:
  + позволяет создавать/редактировать скрипты и пакеты;
  + взаимодействует с другими языками, такими: C, Java и Python;
  + может работать с форматами данных для SAS, SPSS и STATA;
* активное сообщество пользователей и разработчиков;
* регулярные обновления, хорошая документация и тех. поддержка.

Недостатки:

* небольшой объем информации на русском языке;
* относительная сложность в использовании для пользователя, незнакомого с языками программирования.

Среди существующих графических оболочек для работы с R следует отметить платформу R-Studio, которая отличается удобством и постоянно расширяющимися функциональными возможностями.

Главные преимущества платформы R-Studio: высокая гибкость и свободный код. Гибкость позволяет создавать приложения (пакеты) практически на любые потребности. Свободный код даёт возможность разобраться, как именно происходит анализ, а если в коде встретилась ошибка – самостоятельно исправить ее и сделать исправление доступным для всех [10][11].

## Python

Одним из наиболее важных факторов, определяющих популярность Python как языка статистического моделирования, является его широкое использование в качестве предпочтительного языка в науке о данных и машинном обучении. Сегодня существует огромный спрос на знания в области науки о данных, поскольку все больше и больше компаний применяют их в своей деятельности. Python предлагает правильное сочетание мощности, универсальности и поддержки со стороны своего сообщества, чтобы стать лидером.

Специалисты по обработке данных выбирают Python в качестве предпочтительного языка программирования по ряду причин, в том числе:

* ‌открытость исходного кода и активное сообщество;
* ‌более короткая кривая обучения и интуитивно понятный синтаксис;
* ‌большая коллекция мощных и стандартизированных библиотек;
* ‌мощная интеграция с быстрыми компилируемыми языками (например, C / C ++) для примитивов числовых вычислений (используемых в NumPy и pandas);
* ‌простота интеграции основного процесса моделирования с доступом к базе данных, постобработка, такая как визуализация и веб-обслуживание;
* ‌доступность и постоянная разработка интерфейсов Pythonic для фреймворков больших данных, таких как Apache Spark или MongoDB;
* ‌поддержка и разработка библиотек Python крупными и влиятельными организациями, такими как Google или Facebook (например, TensorFlow и PyTorch).

NumPy - это де-факто стандарт численных вычислений в Python, используемый в качестве основы для создания более продвинутых библиотек для приложений анализа данных и машинного обучения, таких как TensorFlow или Scikit-learn. Для числовой обработки NumPy намного быстрее, чем собственный код Python из-за векторизованной реализации его методов и того факта, что многие из его основных подпрограмм написаны на C (на основе инфраструктуры CPython). Он предлагает достойный набор функций статистического моделирования для выполнения базовой описательной статистики и генерации случайных величин на основе различных дискретных и непрерывных распределений.

Matplotlib - это наиболее широко используемая базовая библиотека в Python для общей визуализации.

Seaborn - еще одна мощная библиотека Python, которая построена на основе Matplotlib и предоставляет прямые API-интерфейсы для специальной статистической визуализации и поэтому является фаворитом среди специалистов по данным. Вот некоторые из графиков расширенного статистического моделирования, которые может построить Сиборн:

* ‌тепловые карты;
* ‌сюжеты для скрипки;
* ‌диаграммы рассеяния с линейной регрессией, подгонкой и доверительными интервалами;
* ‌парные графики и графики корреляции, показывающие взаимную зависимость между всеми переменными в таблице данных (с несколькими строками и столбцами);
* ‌графики с фасетами (т.е. визуализация отношений между двумя переменными, которые зависят от более чем одной другой переменной).

SciPy (произносится как «Sigh Pie») –это «экосистема программного обеспечения с открытым исходным кодом на основе Python для математики, естественных наук и инженерии». С помощью SciPy можно делать следующее:

* ‌генерация случайных величин из широкого выбора дискретных и непрерывных статистических распределений - биномиальных, нормальных, бета, гамма, t Стьюдента и т.д;
* ‌частота вычислений и сводная статистика многомерных наборов данных;
* ‌популярные статистические тесты, такие как t-критерий, хи-квадрат, Колмогорова-Смирнова, ранговый критерий Манна-Уитни, ранговую сумму Вилкоксона и т.д;
* ‌корреляционные вычисления, такие как коэффициент Пирсона, дисперсионный анализ, оценка Тейла-Сена и т.д;
* ‌статистические меры расстояния, такие как расстояние Вассерштейна и энергетическое расстояние [12].

## Orange

Orange – это инструмент для визуализации и анализа данных с открытым исходным кодом. Является библиотекой Python. Интеллектуальный анализ данных (Data mining) осуществляется с помощью визуального программирования или сценариев Python. Сценарии Python могут выполняться в окне терминала, интегрированных средах, таких как PyCharm и PythonWin, или оболочках, таких как iPython.

Преимущества Orange для машинного обучения и анализа данных:

* для всех — начинающих и профессионалов;
* выполнить простой и сложный анализ данных;
* создавайте красивую и интересную графику;
* использование в лекции анализа данных;
* доступ к внешним функциям для расширенного анализа.

Процесс построения workflow в Orange происходит путём манипуляций с иконками-виджетами, которые мышкой выкладываются на холст – рабочий стол приложения. Каждый виджет представляет собой программный блок, который каким-либо образом обрабатывает поступившую на его вход информацию и передаёт её дальше, для обработки, визуализации или сохранения следующим виджетом. Связи между виджетами протягиваются мышкой, двойной щелчок открывает окно его настроек: например, отображаемые оси и масштаб для графика и сам график, гиперпараметры для алгоритма машинного обучения, имя файла для виджета загрузки или сохранения данных и т.д. и т.п.

В левой части окна Orange находится блок меню для выбора виджетов. Изначально они сгруппированы в пять разделов:

* Data: виджеты для ввода/вывода данных, фильтрации, выделения и манипулирования выборками, а также (sic!) – большое количество учебных наборов данных;
* Visualize: виджеты для общей (прямоугольная диаграмма, гистограммы, точечная диаграмма) и многомерной визуализации (мозаичная диаграмма, диаграмма-сито);
* Model: набор алгоритмов машинного обучения для классификации и регрессии;
* Evaluate: кросс-валидация, процедуры на основе выборки, оценка методов предсказания;
* Unsupervised: алгоритмы кластеризации (k-средние, иерархическая кластеризация) и проекции данных (многомерное масштабирование, анализ главных компонент, анализ соответствия).

В комплекте начальной установки Orange не содержит, но при необходимости даёт возможность дополнительно загрузить ещё несколько наборов виджетов:

* Associate: датамайнинг повторяющихся наборов элементов и обучение ассоциативным правилам;
* Bioinformatica: анализ наборов генов и доступ к библиотекам геномов;
* Data fusion: объединение различных наборов данных, коллективная матричная факторизация и исследование скрытых факторов;
* Educational: обучение концепциям machine learning;
* Geo: работа с геоданными;
* Image analytics: работа с изображениями, анализ нейронными сетями;
* Network: графовый и сетевой анализ;
* Text mining: обработка естественного языка и анализ текста;
* Time series: анализ и моделирование временных рядов;
* Spectroscopy: анализ и визуализация спектральных наборов данных.

А если и этого недостаточно, то у Orange есть виджет для окончательного решения всех вопросов — Python Script, который позволяет вам написать на Python любой обработчик входных данных [13].

## Loginom

Loginom – low-code платформа для продвинутой аналитики, позволяющая минимизировать написание кода. Платформа содержит обширный набор инструментов анализа и исследования – начиная с простых математических операций и заканчивая нейросетями, что позволяет выстраивать сквозной процесс обработки данных: от ETL до интеллектуального анализа данных и машинного обучения. Предназначен для решения широкого спектра бизнес-задач. Помимо мультизадачности платформа обладает целым рядом конкурентных преимуществ:

* Low-code архитектура подразумевает визуальное проектирование сложных сценариев анализа данных, без привлечения ИТ-департамента и необходимости программирования, силами непосредственно бизнес-пользователя;
* Более 60-и готовых компонентов обработки, как тривиальных, так и достаточно сложных. Например, основанных на алгоритмах машинного обучения (Machine Learning);
* Простая интеграция с различными источниками: файлы (Excel, CSV, XML, Loginom Data File), базы данных (Oracle, MS SQL, PostgreSQL, ClickHouse, BigQuery, MySQL...), бизнес-приложения (1С:Предприятие, Tableau), ODBC-источники (Hive, MongoDB, Amazon Redshift, SQL Azure...), веб-сервисы (SOAP, REST), хранилища данных (Oracle, MS SQL, Firebird);
* Быстрая обработка больших данных (Big Data). Реализована за счет вычислений In-Memory, параллелизма (Parallel Computing);
* Высокая производительность — скорость работы платформы существенно превышает известные на рынке аналоги;
* Переиспользование компонентов позволяет накапливать бизнес-экспертизу. Достаточно один раз создать сценарий обработки данных и сохранить его как готовый компонент, чтобы впоследствии многократно использовать его на новых вводных. Авторские компоненты можно опубликовать, передать, тиражировать, реализовывать на коммерческой основе;
* Создание и публикация веб-сервисов (SOAP (WSDL), REST (JSON)) за пару кликов;
* Асинхронный пользовательский интерфейс — даже при долгих расчетах платформа откликается на действия пользователя, не блокируется;
* Развитая визуализация — OLAP-куб, таблицы, диаграммы, специализированные инструменты;
* Множество готовых бизнес-решений — позволяют реализовывать прикладные задачи: кредитный конвейер, скоринг, очистка данных, управление запасами, process mining и т.д;
* Развертывание без ограничений — stand-alone (инсталляция на ПК), on-premise (установка на сервер), сloud (публичный облачный сервис);
* Разнообразные способы приобретения лицензии — покупка, подписка, Яндекс.Облако;
* Развитая система e-learning обучения Loginom Skills, программа которой рассчитана на различный уровень пользователей: от начинающего до продвинутого. Изучение Loginom включено в образовательную программу 70+ российских вузов, среди которых такие как МГУ им М.В. Ломоносова, НИУ «Высшая школа экономики», РАНХиГС при Президенте РФ;
* Богатая экосистема бесплатных customer-ресурсов: 450+ статей в вики, 100+ материалов в блоге, 150+ видео-роликов. Подготовлена обширная документация по продукту, множество демо-примеров, сервис вопросов и ответов;
* Аналитическая low-code платформа Loginom — 100% российская разработка, входит в Единый реестр отечественного ПО. Loginom Company — независимая компания, не связанная обязательствами со сторонними организациями. Используемые в Loginom компоненты лицензионно безопасны, все библиотеки с исходными кодами (приобретены или Open Source).

Loginom ориентирован на обработку структурированных, т.е. табличных данных. Благодаря своей высокой производительности платформа может с успехом применяться для обработки больших данных.

Для разработки платформы используются 7 языков программирования: Assembler, C, C++, C#, Pascal, TypeScript, JavaScript. Выбор обоснован требованиями обеспечения высокой производительности и экономного расхода оперативной памяти.

Loginom не требует обязательного наличия базы данных или хранилища. Для аналитической платформы они выступают в качестве источника или приемника данных, откуда можно импортировать записи и загрузить после любого этапа обработки.

Тарифы:

* Настольная редакция для некоммерческого использования: бесплатно.
* Настольная редакция для профессиональной аналитики: 97 тысяч рублей.
* Редакция для развертывания в ин
* фраструктуре Яндекс.Облако: от 23 ₽/час.
* Серверная редакция для рабочих групп до 5 человек: от 1.35 млн ₽.
* Серверная редакция для небольших организаций: от 3.9 млн ₽.
* Серверная редакция для крупных компаний и сложных проектов: по запросу.

Есть бесплатная пробная версия на 3 месяца [14][15].

# РАЗДЕЛ 1: РАБОТА В ПАКЕТЕ ORANGE, ЕГО ВОЗМОЖНОСТИ, РЕЗУЛЬТАТЫ ВЫПОЛНЕНИЯ ЗАДАНИЯ

В ходе технологической практики произошло ознакомление с процессом работы в программе Orange (рис. 1-Х).

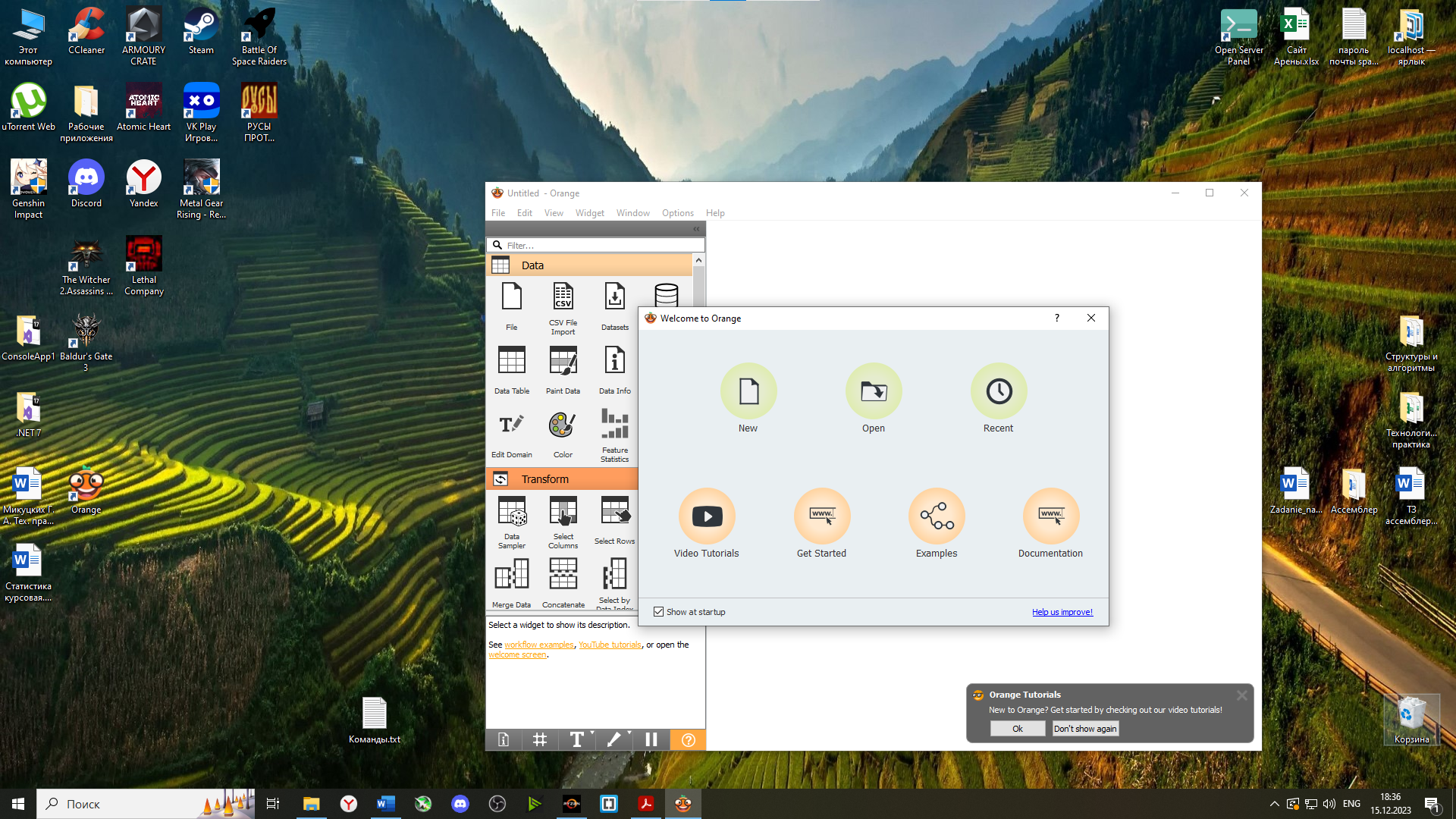


Рис. 1. Создание нового проекта

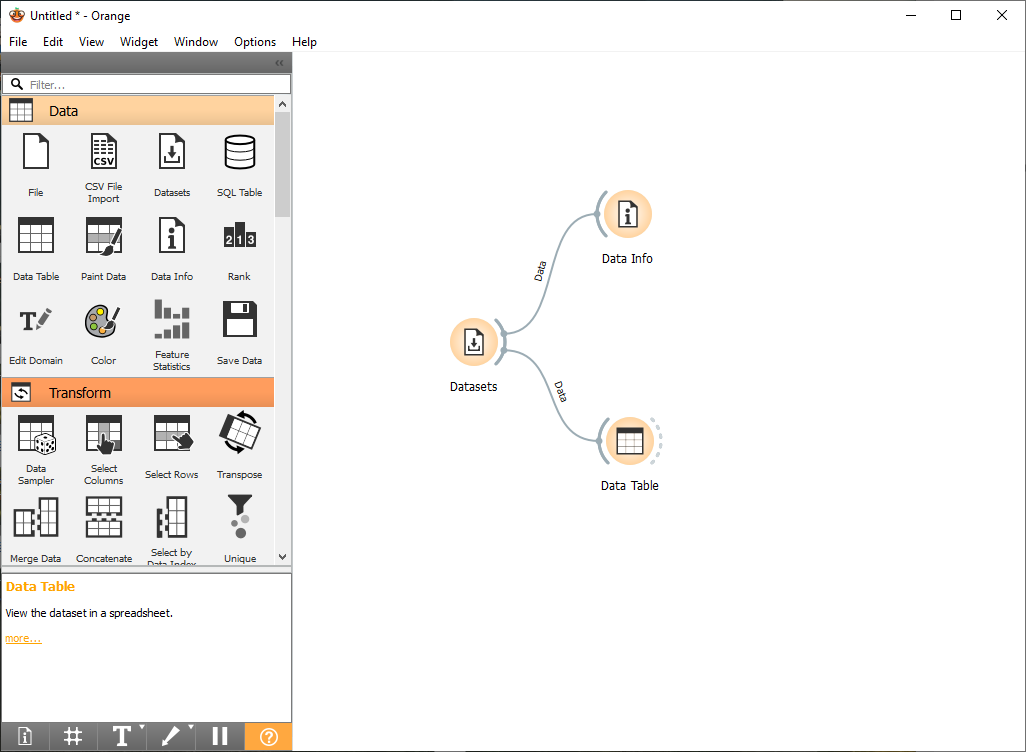


Рис. 2. Подключение встроенного датасета, составление структуры виджетов

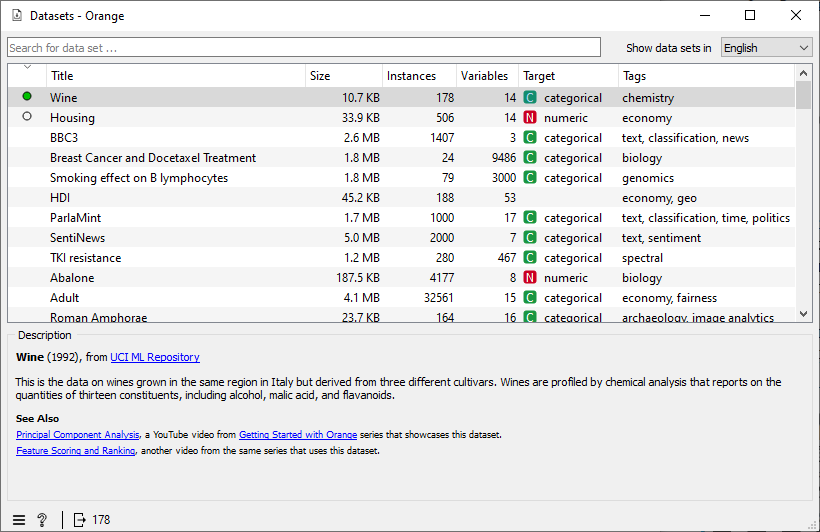


Рис. 3. Меню виджета «Datasets»

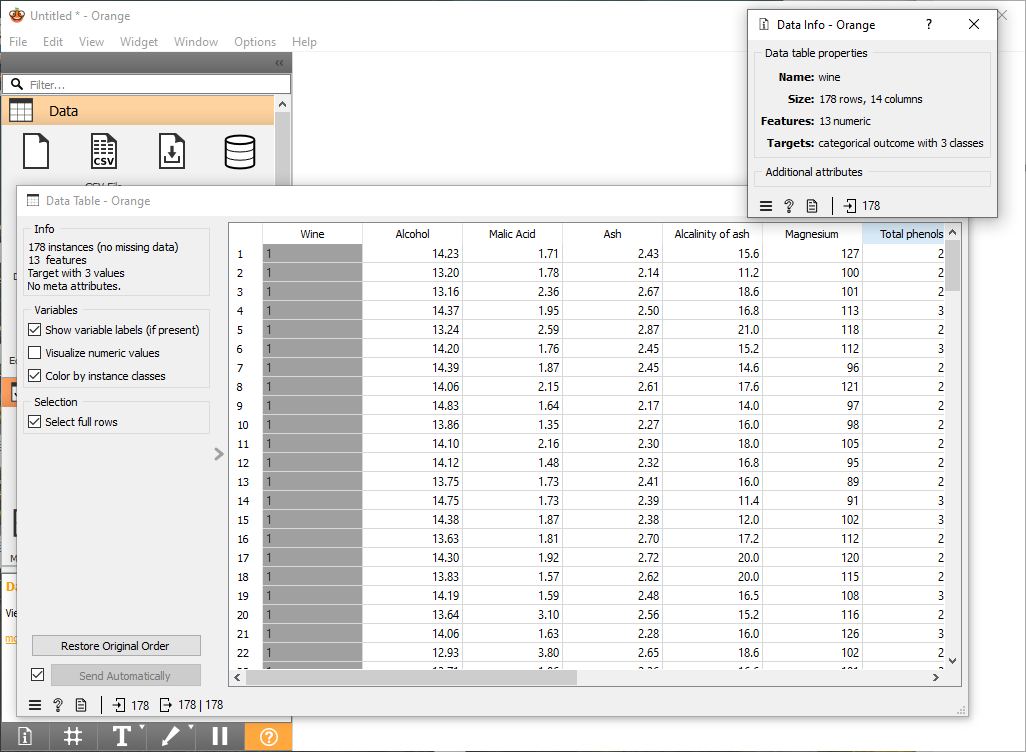


Рис. 4. Меню виджета «Data Table» и «Data Info»

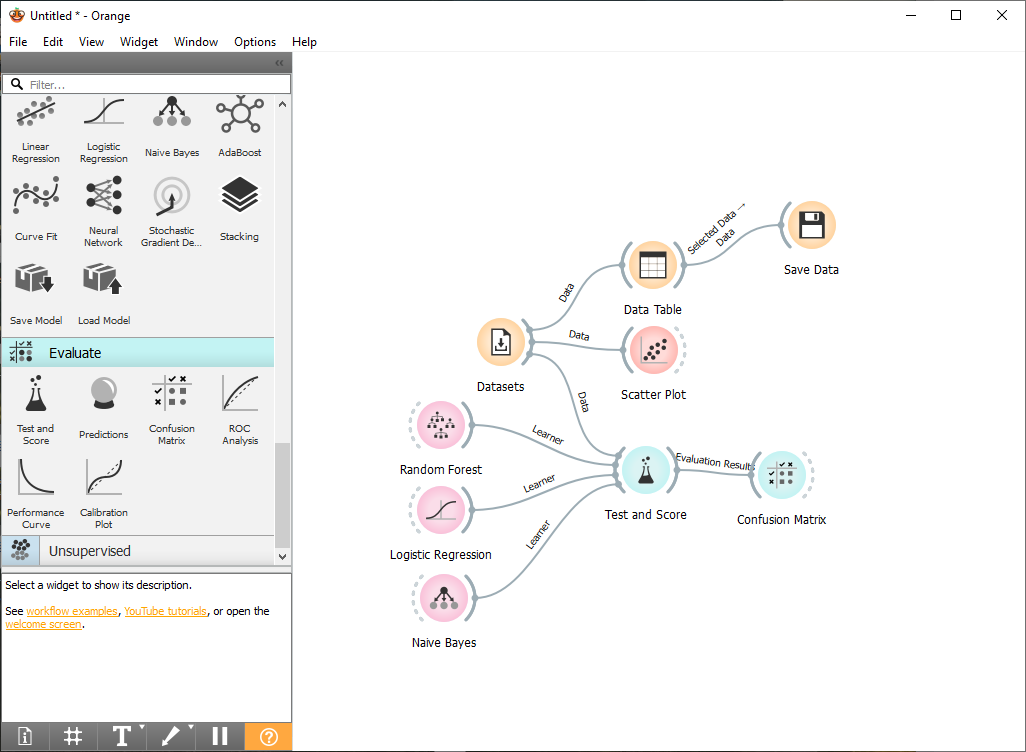


Рис. 5. Создание более сложной структуры обработки данных

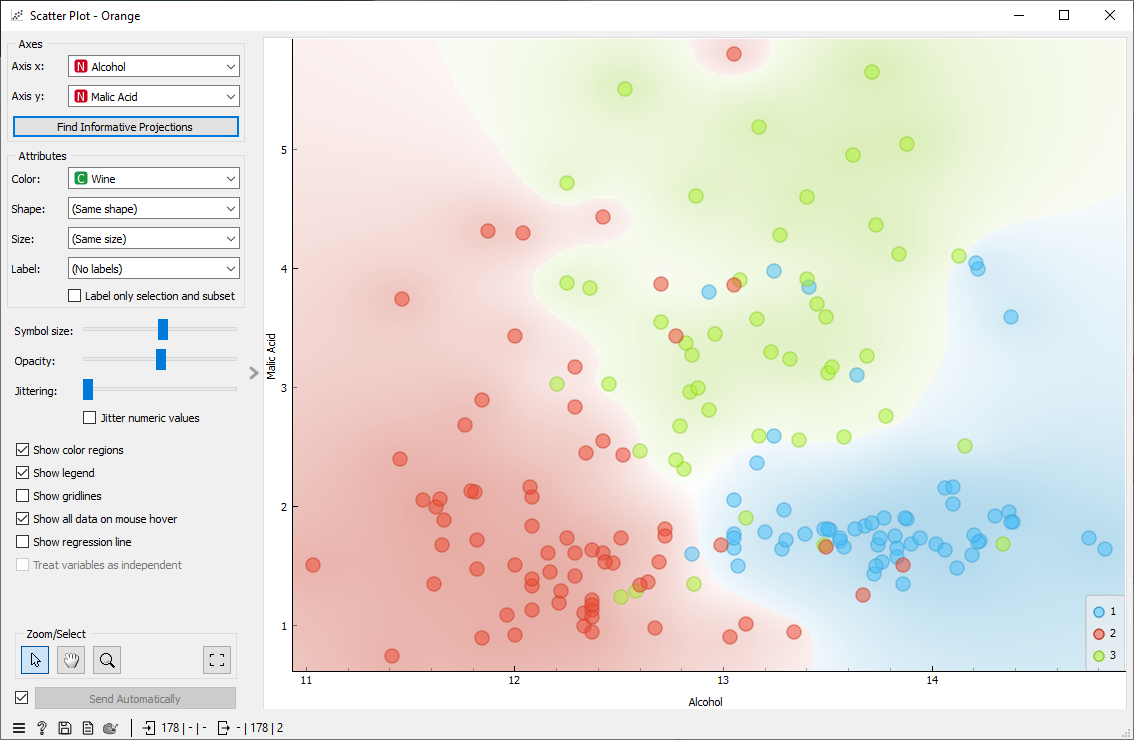


Рис. 6. Меню виджета «Scatter Plot»

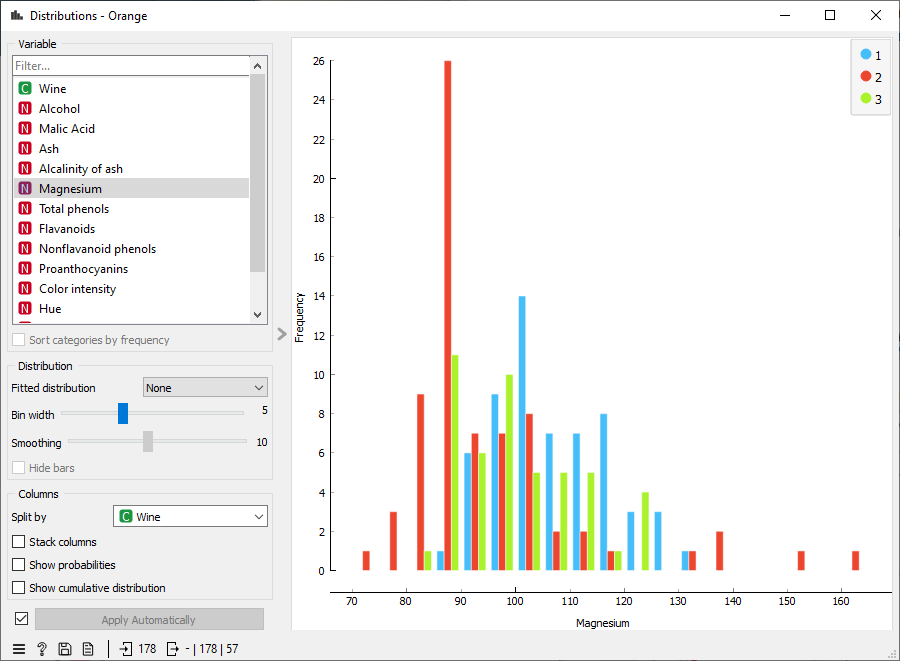


Рис. 7. Меню виджета «Distributions»

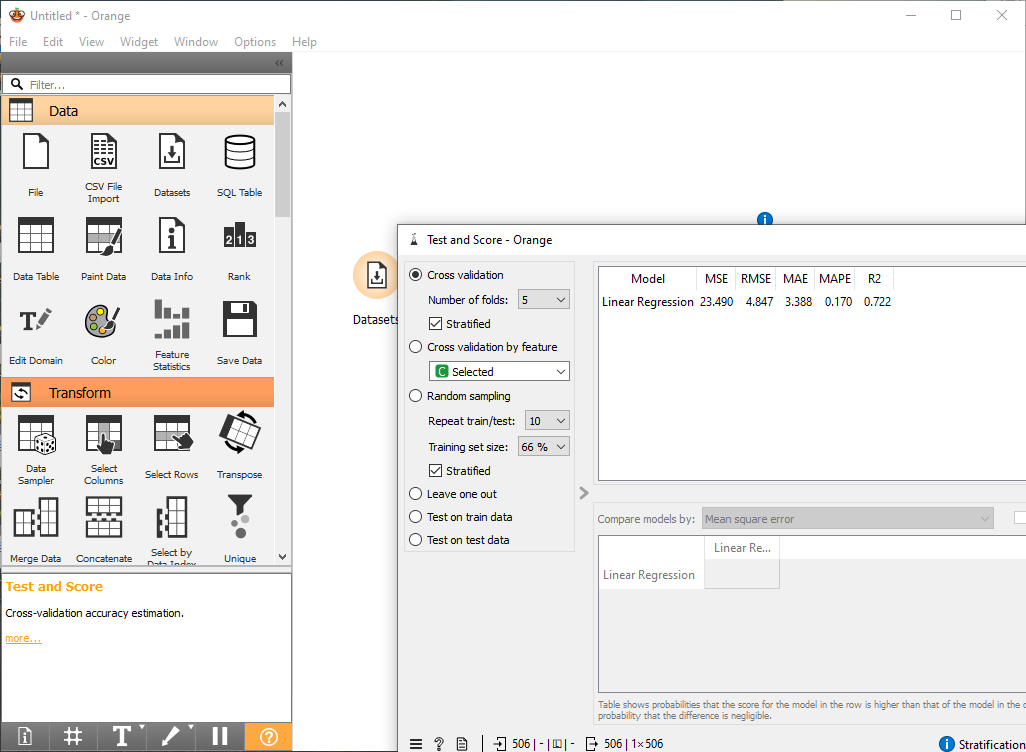


Рис. 8. Виджет «Test and Score» – оценки регрессии

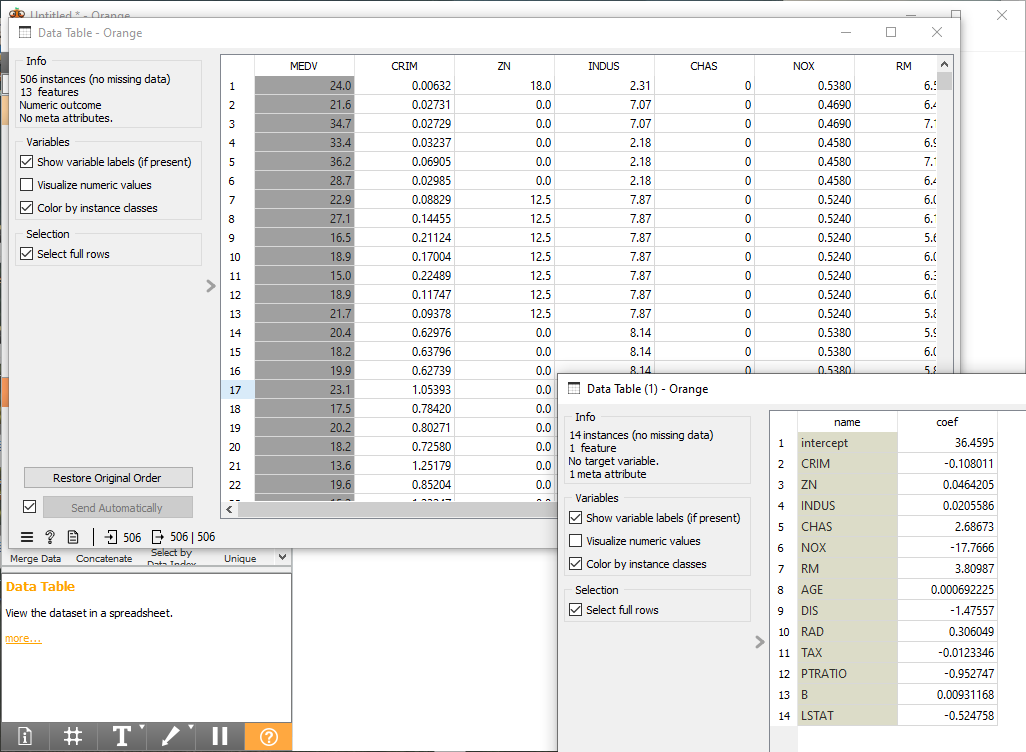


Рис. 9. Таблица регрессии

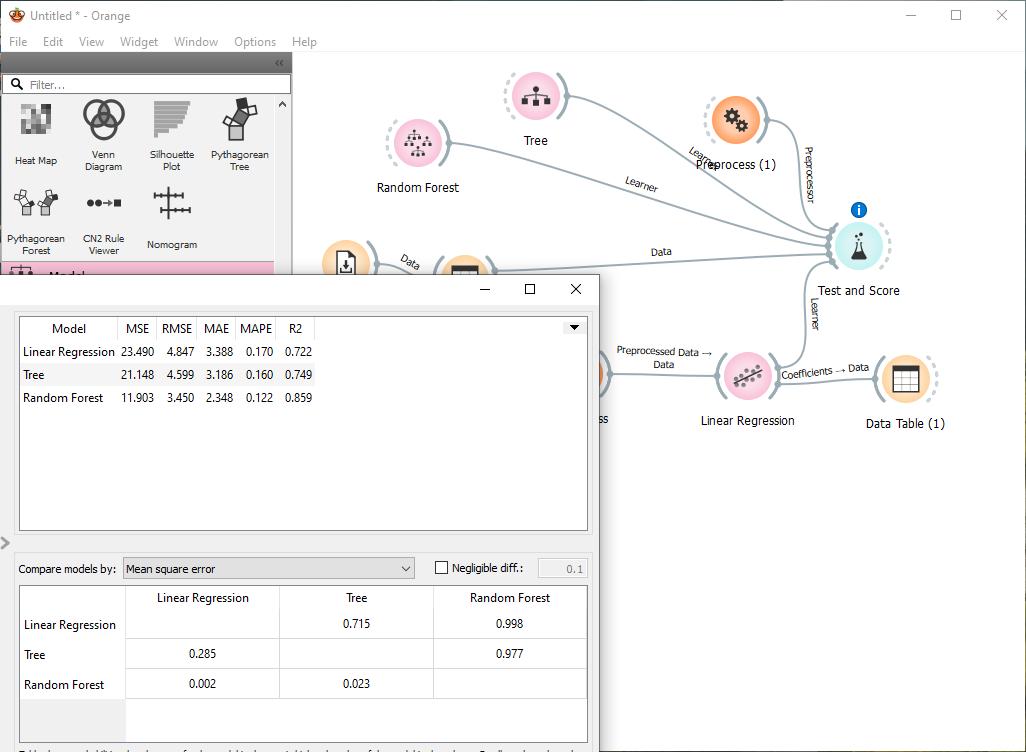


Рис. 10. Анализ коэффициентов регрессии

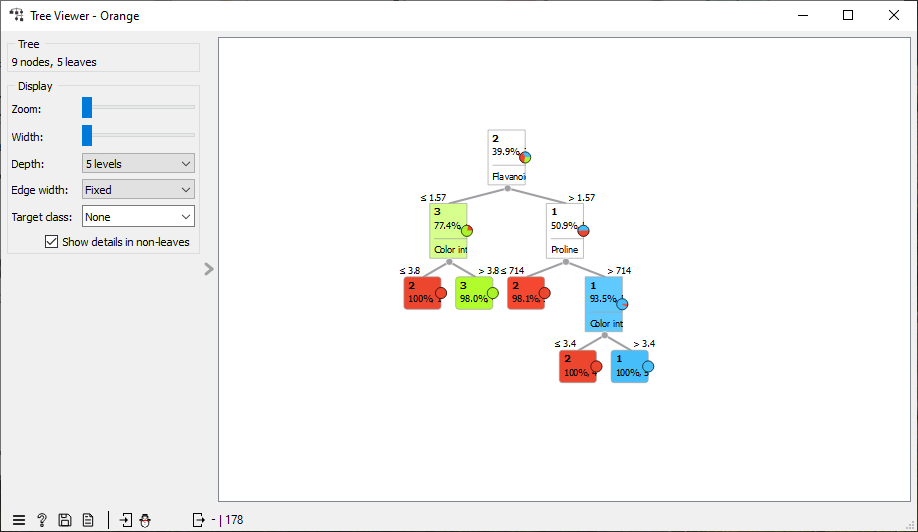


Рис. 11. Построение дерева

# РАЗДЕЛ 2: ОПИСАНИЕ РЕШЕНИЯ ОПТИМИЗАЦИОННОЙ ЗАДАЧИ В MICROSOFT EXCEL

13 вариант

Задание 1: для изготовления трёх видов изделий Р1, Р2, Р3 используют три вида материалов: S1, S2, S3. Составить план выпуска изделий, обеспечивающих их максимальный выпуск по стоимости (рис. 12).



Рис. 12. Решение задания 1

Задание 2: требуется определить минимальную по стоимости смесь сырья для изготовления пищевых концентратов, которые должны содержать питательные вещества (П). Эти вещества содержатся в сырье (М) (рис. 13).



Рис. 13. Решение задания 2

# ЗАКЛЮЧЕНИЕ

В ходе данной технологической практики произошло ознакомление с инструментами для работы с данными, организацией работы в программе Orange, решены задачи линейного программирования в программе Excel с использованием пакета анализа.

В процессе выполнения заданий по ознакомительной практике частично были освоены компетенции:

* способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности (ОПК-2);
* способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности (ОПК-3).

# СПИСОК ЛИТЕРАТУРЫ

[1] https://habr.com/ru/articles/320500/

[2] https://questu.ru/articles/272435/

[3] https://nauchniestati.ru/spravka/intellektualnyj-analiz-onlajn-dannyh-s-pomoshhyu-knime-analytics-platform/

[4] https://nauchniestati.ru/spravka/programma-spss-statistics/

[5] https://studopedia.ru/5\_131347\_paket-SAS.html

[6] https://pandia.ru/text/80/550/1650.php

[7] https://basissoft.ru/min\_product\_mtb.html

[8] https://ru.art-yakin.com/33580-review-of-minitab-fifteen-for-six-sigma

[9] http://www.quality-life.ru/metodologiya01.php

[10] https://handbook.mathpsy.com/?page\_id=1130

[11] https://samoedd.com/soft/r-introduction

[12] https://skine.ru/articles/324190/

[13] https://pro100blogger.com/2019/06/yntellektualnyj-analyz-dannyh-yspolzuem-orange.html

[14] https://loginom.ru/blog/loginom-technique

[15] https://www.osp.ru/articles/2022/1004/13056340