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 27

Студента Богданова Рената Алексеевича группы 1ПИб-02-3оп-23

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. \_\_\_ Виноградова Л.Н.

(должность) (подпись, Ф.И.О.)

2024 год

Содержание

[Введение 3](#_Toc185446660)

[Описание и возможности пакетов для работы с данными 5](#_Toc185446661)

[Раздел 1. Работа в пакете Orange. Его возможности, результаты выполнения задания 12](#_Toc185446662)

[Раздел 2. Описание решения оптимизационной задачи в Excel 21](#_Toc185446663)

[Заключение 26](#_Toc185446664)

[Список литературы 27](#_Toc185446665)

# Введение

В условиях стремительного роста объемов данных в различных сферах деятельности, обработка и анализ информации приобретают ключевую роль. Современные инструменты и методы анализа данных позволяют решать сложные задачи в науке, бизнесе, медицине и других областях. Данная учебная практика направлена на знакомство с современными пакетами для анализа данных и приобретение практических навыков работы с ними.

Учебная практика является обязательным видом учебной работы бакалавра, состоит из двух видов практики: ознакомительная и технологическая.

Цели учебной практики: расширение и закрепление теоретических и практических знаний, полученных студентами при изучении общепрофессиональных и специальных дисциплин; приобретение необходимых умений, навыков и опыта практической работы по изучаемой специальности.

Задачи учебной практики: приобретение практических умений и навыков поиска необходимой для решения задачи информации; формирование профессиональных умений и навыков использования специализированного прикладного программного обеспечения для решения практических задач.

Сроки проведения учебной практики: 2.10.2024-18.12.2024.

В процессе практики бакалавры изучают основы работы со специализированным прикладным программным обеспечением; осуществляют поиск, анализ информации, необходимой для решения поставленных задач; приобретают умения и навыки необходимые для дальнейшей успешной профессиональной деятельности.

Учебная практика базируется на модулях: Математика, Прикладные аспекты математики и физики, Информатика, Программирование.

Учебная практика направлена на закрепление связи теоретического обучения с практической деятельностью, на повышение уровня практической подготовки студентов, обеспечивает прикладную направленность и адаптацию к рынку труда.

Результаты, полученные в процессе выполнения заданий практики должны стать основой для понимания таких модулей как: Инструментальные средства программирования, Архитектура программно-информационных систем, Проектирование систем управления данными, Технология разработки программного обеспечения, а также для производственной практики.

# Описание и возможности пакетов для работы с данными

1. KNIME

KNIME (Konstanz Information Miner) [1] — это одна из наиболее востребованных и эффективных платформ для анализа данных и машинного обучения. Она предлагает удобный графический интерфейс, который позволяет пользователям создавать и выполнять рабочие процессы (пайплайны) без необходимости писать код. Данная платформа предназначена для комплексного решения задач в области науки о данных, от создания аналитических моделей до их развертывания, обмена информацией внутри организации, а также создания приложений и сервисов на основе данных.

KNIME предлагает как бесплатный, так и платный варианты использования [2]. Бесплатная версия KNIME, Personal Plan, доступна для индивидуальных пользователей и позволяет просматривать, загружать и обмениваться рабочими процессами и компонентами с сообществом, хранить рабочие процессы в личных пространствах, сотрудничать с другими пользователями и отслеживать историю версий рабочих процессов и компонентов.

Для успешного применения KNIME нужны основы науки о данных и машинного обучения, понимание типов данных и форматов файлов, умение обрабатывать и преобразовывать данные. Знание Python или R будет полезным, но не обязательным. Также важно уметь визуализировать и интерпретировать данные, понимать статистические методы и аналитические модели.

KNIME не требует умения программировать, но знание Python или R позволит расширить функциональность платформы.

KNIME работает с различными типами данных (структурированными, неструктурированными и полуструктурированными) и поддерживает подключение к базам данных, файлам, web-сервисам и социальным сетям.

Результатами работы KNIME могут быть интерактивные отчеты и визуализации данных, статистические модели и результаты анализа, машинно-обученные модели и прогнозы, автоматизированные рабочие процессы анализа данных, приложения и сервисы на основе данных.

1. SPSS

SPSS (Statistical Package for the Social Sciences) [3] - это платный коммерческий статистический пакет, широко используемый в социальных науках, а также в здравоохранении, маркетинге и образовании. Для студентов и преподавателей доступны академические лицензии.

SPSS позволяет проводить статистический анализ данных, вводить и хранить данные разных типов, а также предоставляет широкий набор инструментов для проведения различных статистических расчетов. Программа обладает удобным графическим интерфейсом, работает с собственным форматом данных (.sav) и импортирует данные из других форматов, например, текстовых файлов, таблиц Excel и баз данных.

Результатами работы SPSS являются детальные отчеты с результатами анализа, включая таблицы, графики и статистические показатели, а также модели для прогнозирования и анализа данных.

Знание языка программирования SPSS позволяет автоматизировать задачи и расширить функциональность программы, но не является обязательным.

1. SAS

SAS [4] — комплексная система для статистического анализа данных, разработанная компанией SAS Institute Inc. Она распространяется на платной основе, но для обучения доступна бесплатная версия SAS University Edition.

SAS предназначен для широкого спектра задач, связанных с анализом данных: обработка, анализ, визуализация, моделирование. Он позволяет проводить математические и статистические расчеты, строить модели, создавать отчеты и анализировать данные различных форматов.

Система предлагает набор инструментов для работы с данными: импорт, очистка, преобразование, агрегирование, визуализация, статистический анализ, прогнозирование и моделирование. Для успешного использования SAS необходимо знание основ работы с данными и навыков программирования на языке SAS. Хотя для некоторых задач достаточно визуального интерфейса SAS Enterprise Guide, для более сложных задач требуется написание программного кода. SAS работает с различными форматами данных: текстовые файлы, таблицы, базы данных и другие источники.

Результатом работы с SAS может быть получение ответов на вопросы, выявление закономерностей и трендов, построение прогнозных моделей, создание отчетов и презентаций, разработка решений для принятия решений в различных сферах деятельности.

1. STATISTICA 6.0

STATISTICA 6.0 [6] представляет собой коммерческий программный пакет, не имеющий бесплатной версии для академических учреждений. Однако, на сайте производителя доступна пробная версия, позволяющая ознакомиться с функционалом программы. STATISTICA 6.0 предназначена для проведения широкого спектра статистических анализов, включая математические расчеты, моделирование, визуализацию данных, построение гистограмм, диаграмм рассеяния и других графических представлений. Программа предлагает набор инструментов для работы с различными типами данных, включая числовые, текстовые, временные ряды и другие.

Для успешной работы с STATISTICA 6.0 необходимо обладать базовыми знаниями в области статистики и математики. Знание основ программирования не является обязательным, однако, понимание синтаксиса команд и функций позволит оптимизировать работу с программой. STATISTICA 6.0 работает с различными типами данных, в том числе с данными, полученными из табличных редакторов, баз данных и других источников. Результатом работы с программой являются отчеты, содержащие таблицы, графики и результаты статистического анализа.

1. Minitab

Minitab [7] - статистический пакет, доступный в коммерческой и академической версиях. Последняя предоставляется бесплатно студентам и преподавателям. Программа предназначена для решения задач статистики, анализа данных и моделирования.

Minitab предлагает инструменты для математических расчетов, статистического анализа, визуализации и моделирования. Он позволяет проводить различные статистические тесты, строить графики, анализировать временные ряды, контролировать качество, проводить регрессионный и многомерный анализ.

Для успешного применения Minitab необходимы базовые знания статистики и анализа данных. Программирование не является обязательным, но может повысить эффективность работы. Minitab работает с данными в различных форматах, результатом работы являются статистические выводы, аналитические отчеты, графики и модели, которые помогают в принятии решений.

1. R + R-Studio

R [8] – свободно распространяемое программное обеспечение с открытым исходным кодом, доступное для загрузки и использования без лицензионных сборов. Для коммерческих целей доступны платные версии R-Studio с расширенным функционалом. Однако для академических целей достаточно бесплатной версии, предоставляющей весь необходимый функционал для студентов и преподавателей.

R предназначен для статистических вычислений, анализа данных и моделирования. Он позволяет проводить математические расчеты, моделировать процессы, анализировать данные с использованием широкого спектра статистических методов.

R предлагает набор инструментов для работы с данными: статистические тесты, построение графиков, анализ временных рядов, классификацию, регрессионный анализ. Он работает с таблицами, текстовыми файлами, базами данных и web-данными.

Для успешного применения R необходимы знания основ статистики, анализа данных и языка программирования R. Результатом работы с R являются аналитические отчеты, графики, статистические модели и выводы, помогающие в решении аналитических задач.

1. Python

Python [9] - язык программирования с открытым исходным кодом, доступный для бесплатного использования и распространения. Для профессионального применения доступны платные пакеты, но для академических целей студентам, как правило, доступен бесплатный режим. Python широко применяется в научных исследованиях, анализе данных, машинном обучении, web-разработке, автоматизации и обработке текстов.

Он предоставляет инструменты для математических расчетов, моделирования и статистических вычислений, а также обладает широким набором библиотек, расширяющих его возможности. Для успешного применения Python необходимы знания основ программирования, включая типы данных, переменные, операторы, циклы, условия, функции. Дополнительно желательно владение алгоритмами, структурами данных и парадигмами программирования.

Python может работать с различными форматами данных, включая текстовые файлы, таблицы, базы данных, изображения, аудио- и видеофайлы. Результат работы с Python может быть представлен в виде программ, результатов анализа данных, моделей машинного обучения или веб-приложений.

1. Orange

Orange [10] — это бесплатное программное обеспечение с открытым исходным кодом, предназначенное для визуального анализа данных и машинного обучения. Оно предоставляет широкий набор инструментов для проведения математических расчетов, моделирования, статистического анализа и визуализации данных. Orange не требует глубоких навыков программирования, так как работает на основе визуальной среды, позволяющей создавать и запускать рабочие процессы анализа данных без написания кода.

Пользователи могут создавать комплексные рабочие процессы, собирая из виджетов - готовых блоков для загрузки данных, предварительной обработки, визуализации, обучения моделей, прогнозирования и оценки результатов. Для успешного применения Orange требуется базовое понимание принципов анализа данных и машинного обучения.

Программное обеспечение работает с различными типами данных, включая табличные данные, текстовые данные, изображения, аудио и видео. Результатом работы с Orange является получение результатов анализа данных, включая графики, диаграммы, статистические показатели, обученные модели и прогнозы.

1. Loginom

Аналитическая платформа Loginom [11] доступна в нескольких вариантах: бесплатная редакция Community, предназначенная для некоммерческого использования, и платные редакции, предоставляющие расширенный функционал. Для платных пакетов академический режим (бесплатный для студентов) не предусмотрен.

Loginom представляет собой инструмент для визуального анализа данных, позволяющий выполнять как элементарный, так и сложный анализ без написания кода. Платформа обладает широким спектром возможностей, включающим математические расчеты, построение моделей, статистический анализ, визуализацию данных и разработку отчётов. Loginom предоставляет набор инструментов: компоненты для импорта и экспорта данных, визуальные редакторы для построения схем и диаграмм, графические инструменты для визуализации данных, библиотеки статистических функций и машинное обучение.

Для успешного применения платформы необходимы базовые знания основ анализа данных, математической статистики и работы с таблицами. Знание языков программирования не является обязательным, так как Loginom предоставляет визуальный интерфейс для выполнения всех операций. Платформа работает с различными типами данных: табличные базы данных, CSV-файлы, текстовые файлы, а также внешние источники, такие как веб-сервисы. Результатом работы с Loginom является получение аналитической информации в виде графиков, отчётов, прогнозов, моделей и других визуализаций, которые могут быть использованы для принятия решений, оптимизации процессов и повышения эффективности работы.

Вывод по инструментам

Для начинающих аналитиков данных рекомендуются инструменты с визуальным интерфейсом, не требующие навыков программирования, такие как KNIME, Orange и Loginom. Они позволяют создавать и запускать рабочие процессы без необходимости писать код, что делает их доступными для широкого круга пользователей.

Опытные аналитики данных могут отдать предпочтение R + R-Studio и Python, которые предоставляют большую гибкость и мощные возможности, но требуют владения соответствующими языками программирования.

SPSS и SAS широко применяются в узкоспециализированных областях, таких как социальные науки и маркетинг, предлагая специализированные инструменты для решения задач в этих сферах. Их использование требует глубокого понимания соответствующих концепций и методов.

Minitab является удобным инструментом для анализа данных, контроля качества и регрессионного анализа. Он не требует глубокого знания программирования, но его функциональность может быть ограничена.

Ознакомившись с представленными инструментами анализа данных, можно сделать вывод, что выбор инструмента зависит от уровня подготовки пользователя и конкретной задачи.

# Раздел 1. Работа в пакете Orange. Его возможности, результаты выполнения задания

Данный раздел посвящён ознакомительной работе с программой Orange.

Определим ключевые термины, которые помогут в дальнейшем понимании:

* виджет - основная точка обработки любых действий с данными;
* рабочий процесс — это последовательность шагов или действий, которые вы выполняете на платформе для решения задачи.

После установки и запуска программы открывается пустой рабочий процесс (см. рис. 1).

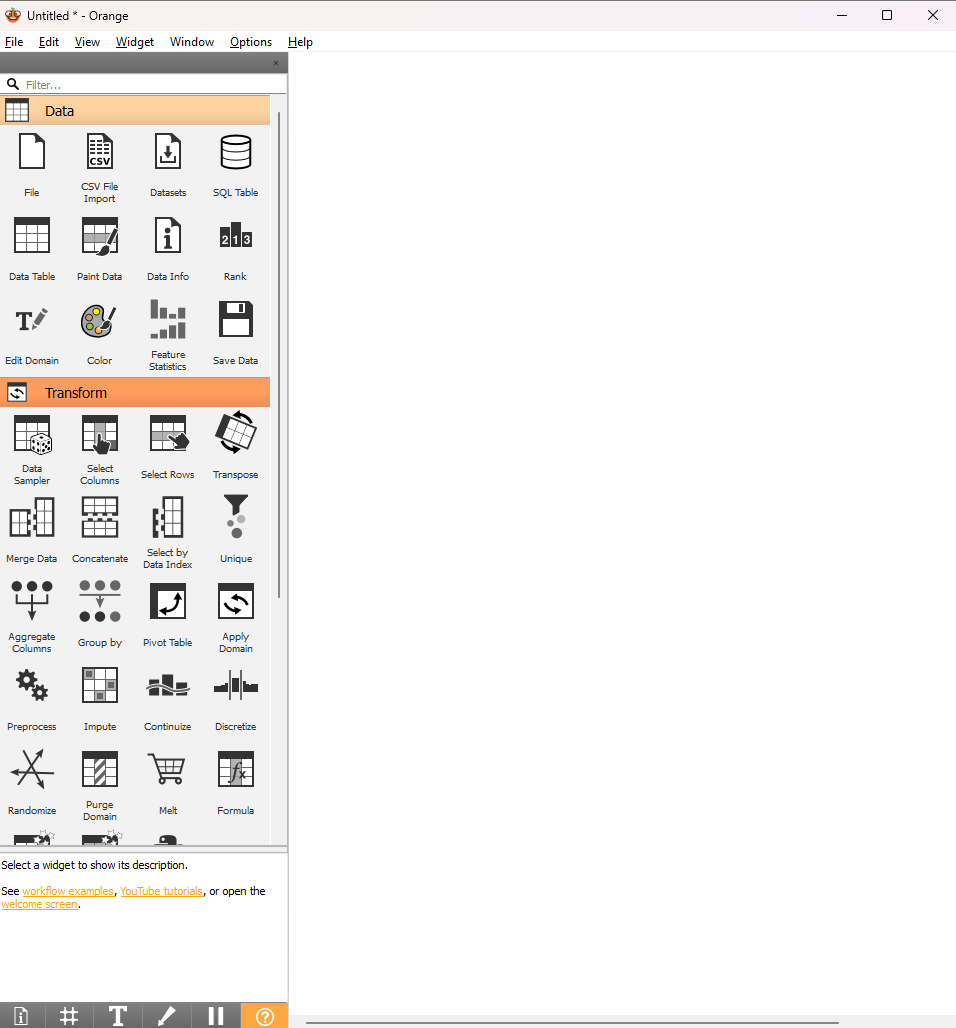


Рис. 1. Пустой рабочий процесс

Далее необходимо импортировать данные. Для этого из меню выбора виджетов перетаскиваем виджет «Datasets» в рабочую область (см. рис. 2).

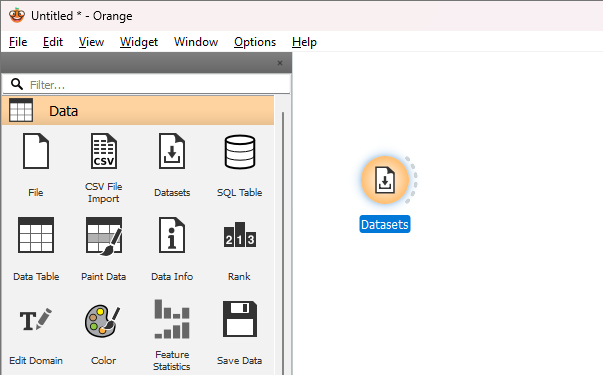


Рис. 2. Виджет «Datasets»

Загрузим данные. Для этого дважды кликаем левой кнопкой мыши по виджету. Открывается таблица, из которой можно выбрать один из встроенных набора данных. Выберем датасет «Kickstarter projects» (см. рис. 3).

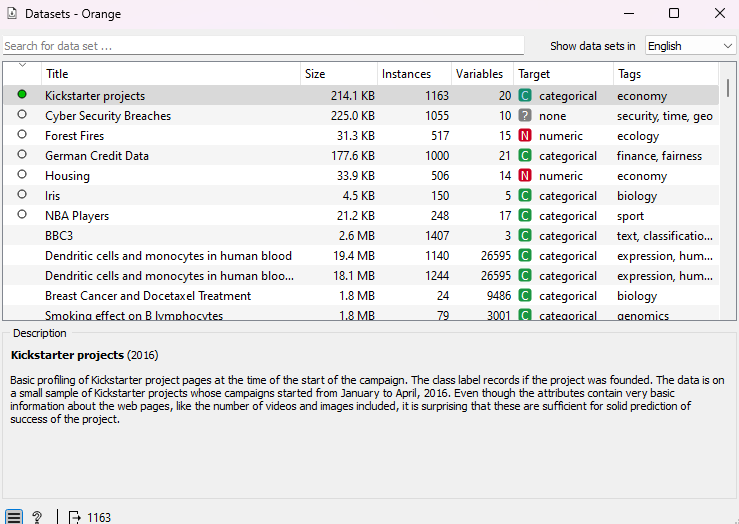


Рис. 3. Выбор набора данных

Перенесём в рабочую область ещё два виджета: «Data info» и «Data Table». Соединим их с виджетом «Datasets» (см. рис.4).

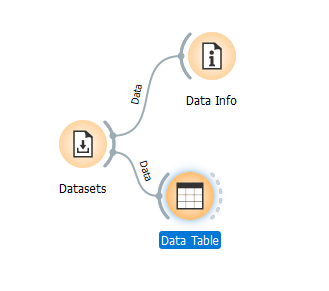


Рис. 4. Виджеты «Data info» и «Data Table»

Виджет «Data info» позволяет посмотреть информацию о наборе данных (см. рис. 5).

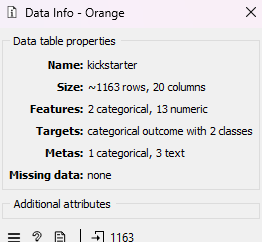


Рис. 5. Информация о наборе данных

«Data Table» даёт возможность посмотреть данные в табличном виде (см. рис. 6).

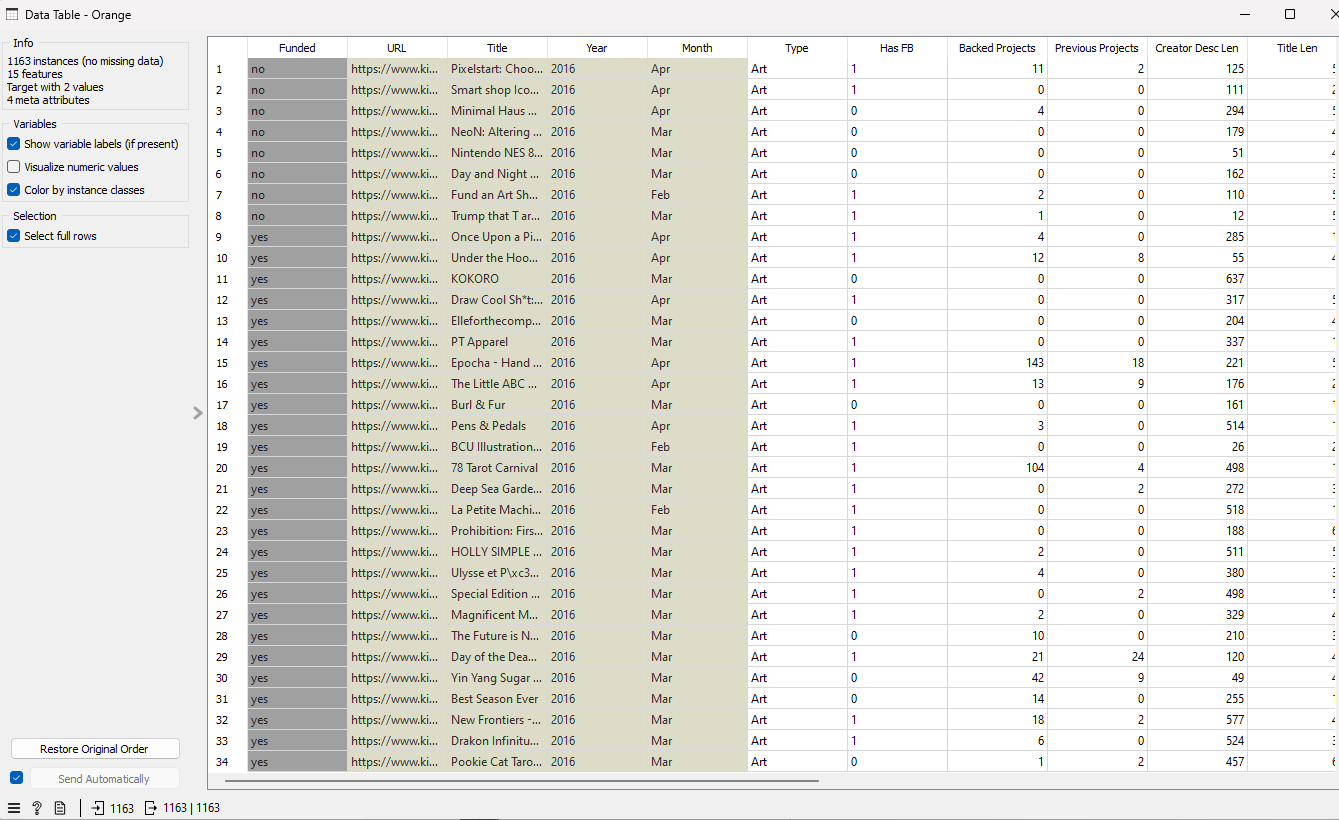


Рис. 6. Меню виджета «Data Table»

Добавим виджет «Scatter Plot» (см. рис. 7). Scatter plot (точечная диаграмма) — это способ представления данных, в котором каждый объект представляется точкой. Каждая такая точка характеризуется двумя величинами: по оси X и по оси Y (см. рис. 8).

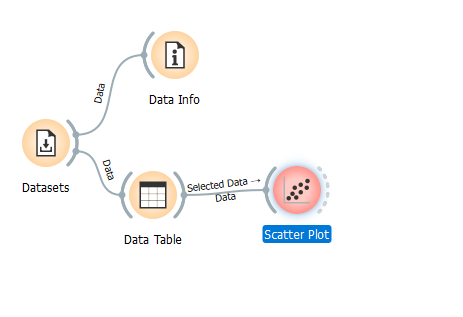


Рис. 7. Виджет «Scatter Plot»

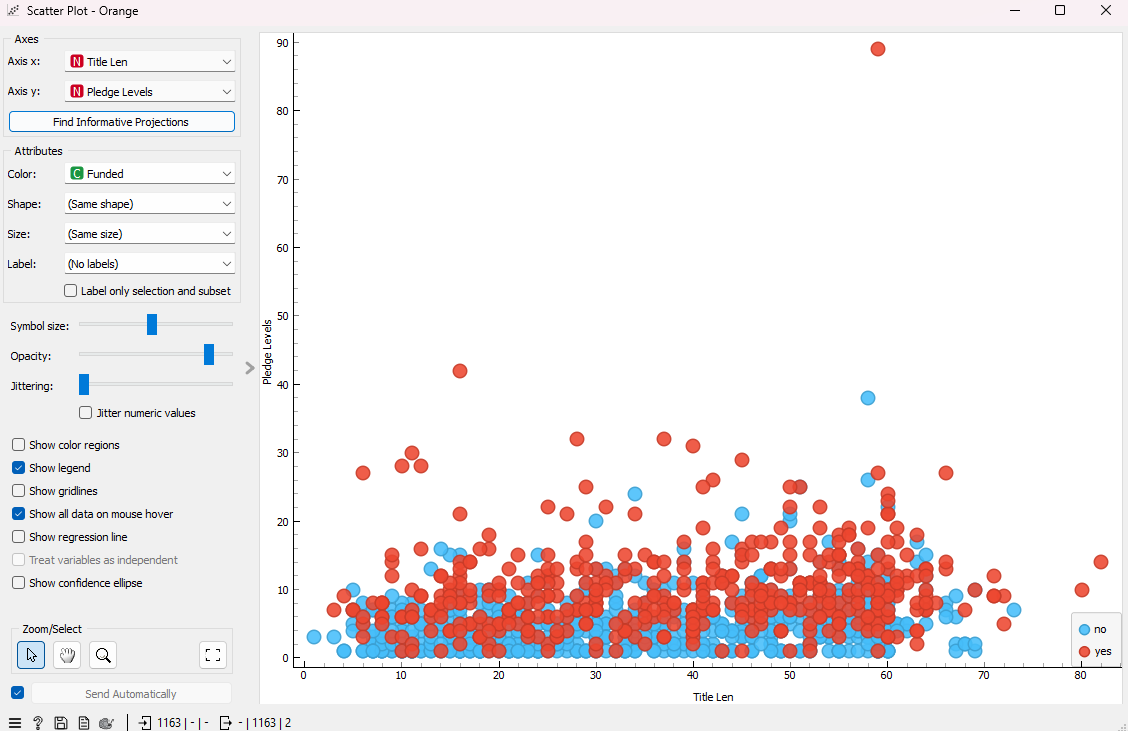


Рис. 8. Пример точечной диаграммы

С помощью виджета «Distributions» можно посмотреть, как распределены числовые переменные (см. рис. 9).

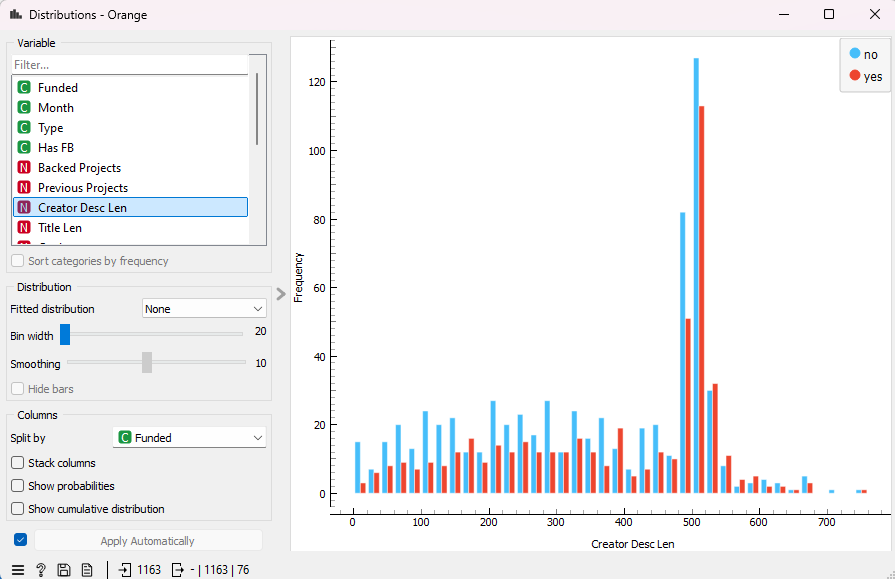


Рис. 9. Виджет «Distributions»

Для того, чтобы построить регрессию, используем виджеты «Select Columns» - для выбора столбцов из таблицы, «Linear Regression» - для настройки регрессии, и «Test and Score» - для оценки регрессии (см. рис. 10).

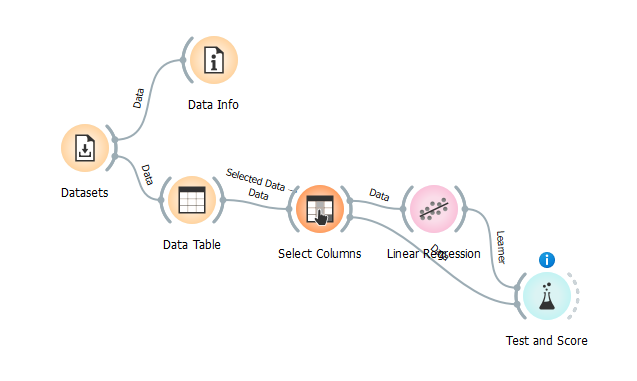


Рис. 10. Виджеты, необходимые для построения регрессии

Для просмотра результатов регрессии нужно кликнуть по виджету «Test and Score». В открывшемся меню мы увидим таблицу с оценками (рис. 11).

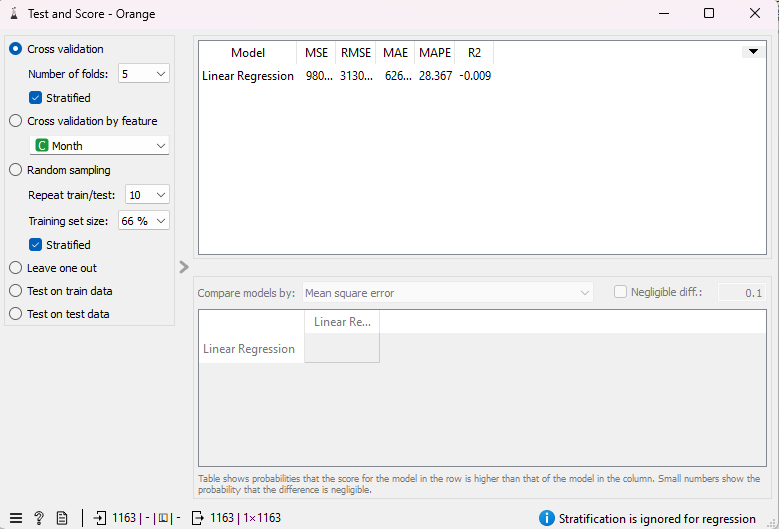


Рис. 11. Оценка регрессии

Проанализируем значения оценок модели, такие как средний квадрат ошибки (MSE), среднеквадратическое отклонение (RMSE), средняя абсолютная ошибка (MAE) и коэффициент детерминации (R²). MAE, будучи линейной оценкой, равномерно взвешивает все различия между истинными значениями и предсказаниями. Для улучшения результатов расчетов и повышения точности модели добавим предварительную обработку данных (препроцессинг) и построим деревья решений (см. рис. 12-13).

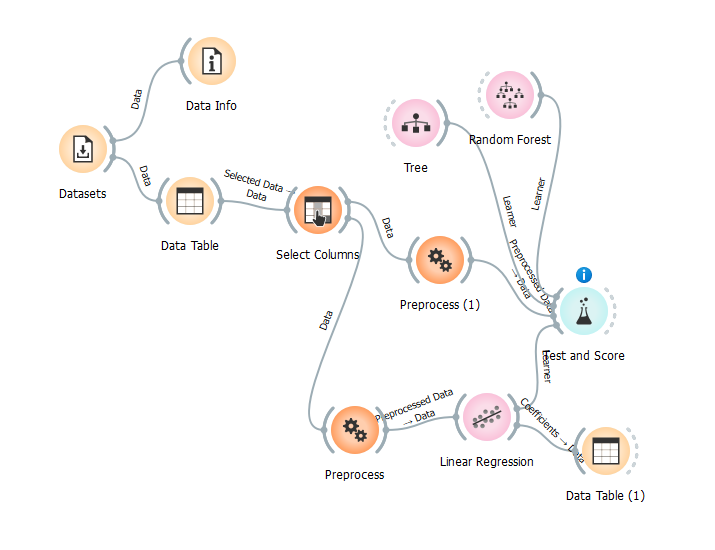


Рис. 12. Добавление виджетов «Preprocess», «Tree» и «Random Forest»

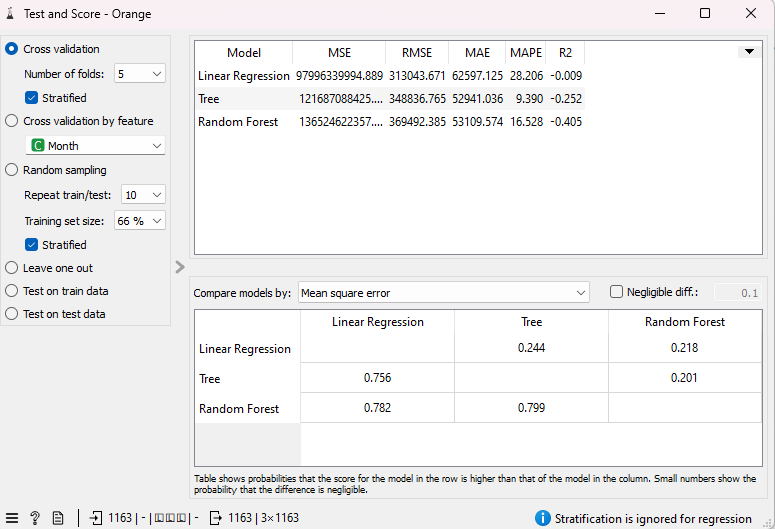


Рис. 13. Оценка регрессии с учётом препроцессинга

Посмотрим на визуализацию дерева с помощью виджета «Tree Viewer» (см. рис. 14-15).

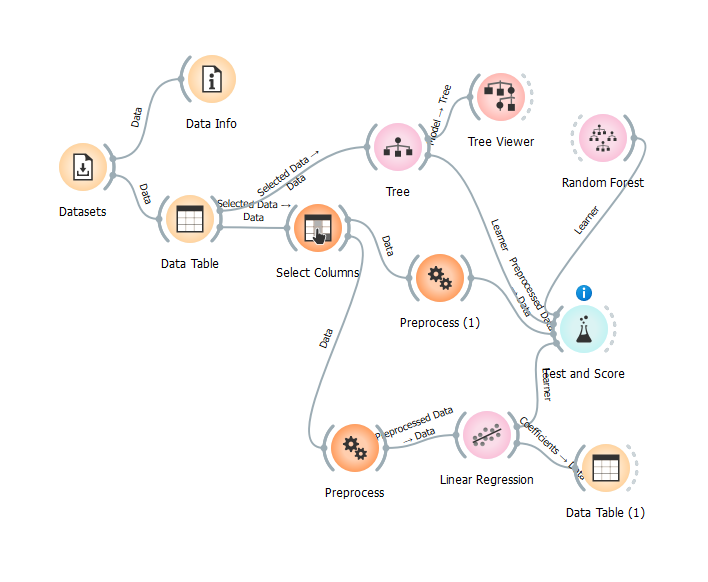


Рис. 14. Добавление виджета «Tree Viewer»

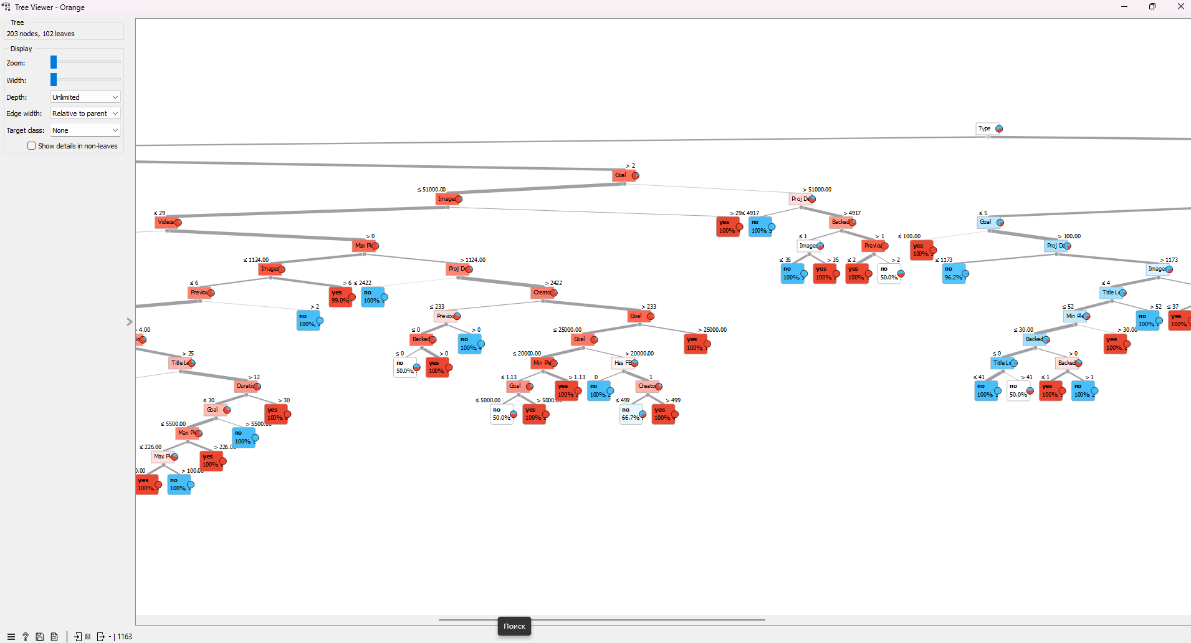


Рис. 15. Фрагмент визуализации дерева

Данный раздел был посвящен ознакомлению с программным пакетом Orange. В ходе работы были изучены основные принципы работы с программой, включая понятие виджетов и рабочих процессов.

Для демонстрации возможностей пакета был использован датасет «Kickstarter projects». Были продемонстрированы основные функции Orange, такие как импорт и визуализация данных, построение линейной регрессии, а также использование препроцессинга для повышения точности модели.

В завершение была рассмотрена возможность построения и визуализации дерева решений, что позволяет получить более глубокое понимание функционирования модели и ее прогнозов.

Полученные результаты демонстрируют, что пакет Orange является мощным инструментом для проведения анализа данных и построения моделей машинного обучения, доступным как для начинающих, так и для опытных пользователей.

# Раздел 2. Описание решения оптимизационной задачи в Excel

Задание 1

Для изготовления трёх видов изделий P1, P2 и P3 используют три вида материалов: S1, S2, S3. Запасы материалов, технологические нормы расхода материалов на каждое изделие и цена единицы изделия приведены в таблице. Составить план выпуска изделий, обеспечивающих их максимальный выпуск.

Вариант задания 19: a11=5, a12=5, a13=2, a21=4, a22=6, a23=8, a31=5, a32=6, a33=2; b1=250, b2=160, b3=530, c1=10, c2=10, c3=12.

Запишем ограничения задачи:

Запишем целевую прибыль:

Для решения задачи составим таблицу (см. рис. 16-17).

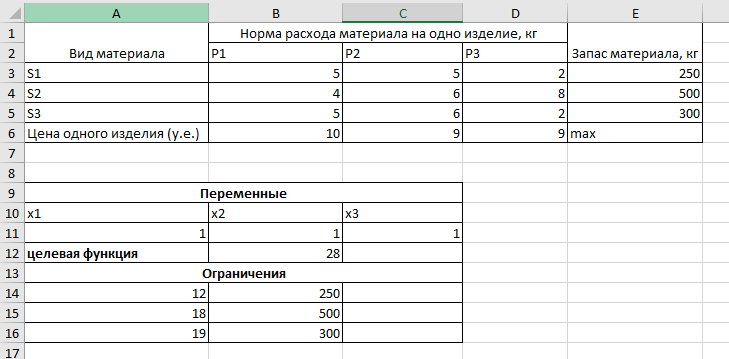


Рис. 16. Таблица для задачи 1 (значения)

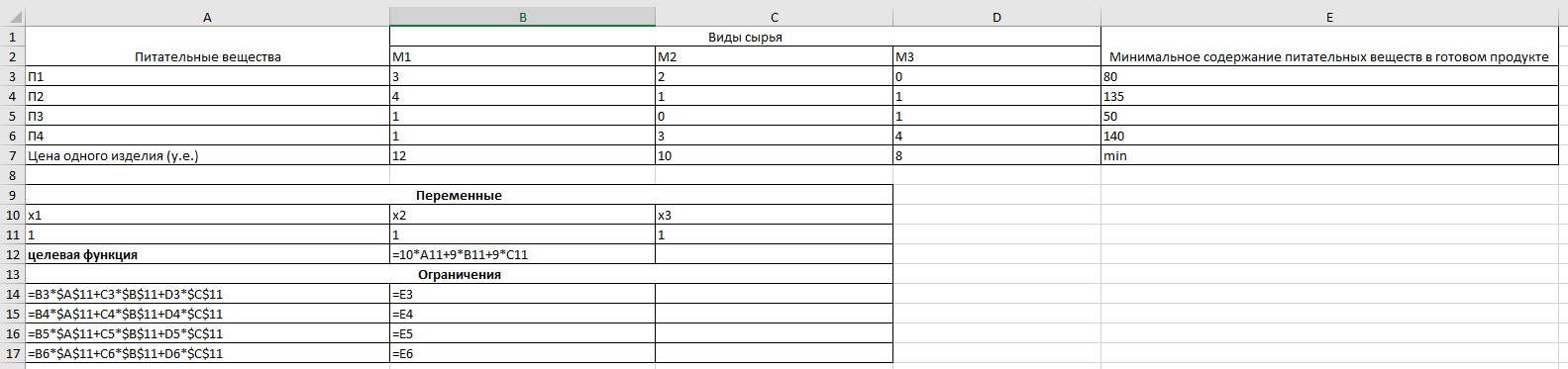


Рис. 17. Таблица для задачи 1 (формулы)

Затем, используя команду “Поиск решения”, заполним открывшееся окно и нажмём кнопку “Найти решение” (см. рис. 18).

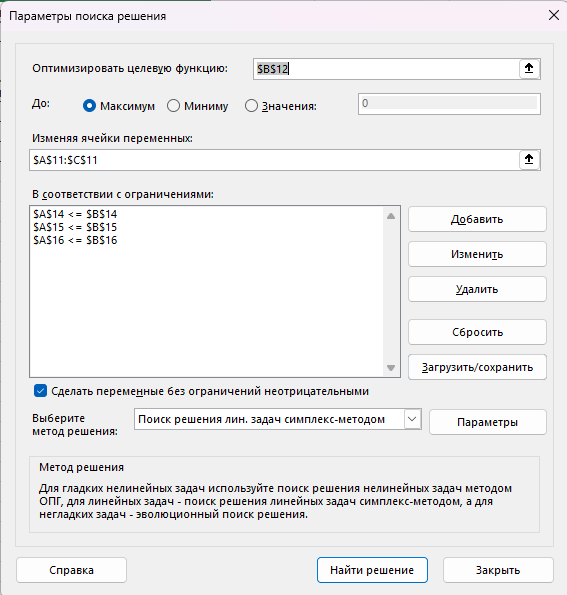


Рис. 18. Поиск решения для задачи 1

После проведения расчётов мы получим результат (см. рис. 19).

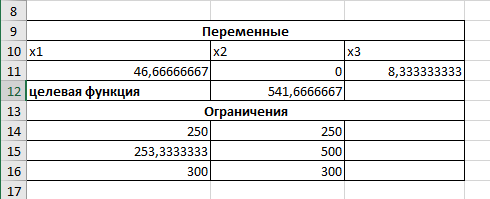


Рис. 19. Результаты для задачи 1

На основе полученных результатов можно сделать следующий вывод: для достижения максимальной прибыли предприятию следует сосредоточиться на производстве изделий P1 (46 шт.) и P3 (8 шт.), поскольку производство изделия P2 не приносит дополнительной прибыли при данных условиях. Это решение учитывает все ограничения по доступным запасам материалов S1, S2 и S3, обеспечивая их полное использование без превышения лимитов.

Задание 2

Требуется определить минимальную по стоимости смесь сырья для изготовления пищевых концентратов, которые должны содержать питательные вещества (П). Эти вещества содержаться в сырье (М) в различных сочетаниях. Содержание питательных веществ в сырье и готовом продукте, а также цена на каждый вид сырья показаны в таблице.

Вариант задания 19: a11=3, a12=2, a13=0, a21=4, a22=1, a23=1, a31=1, a32=0, a33=1, a41=1, a42=3, a43=4; b1=80, b2=135, b3=50, b4=140; c1=12, c2=10, c3=8.

Запишем ограничения задачи:

Запишем целевую прибыль:

Для решения задачи составим таблицу (см. рис. 20).

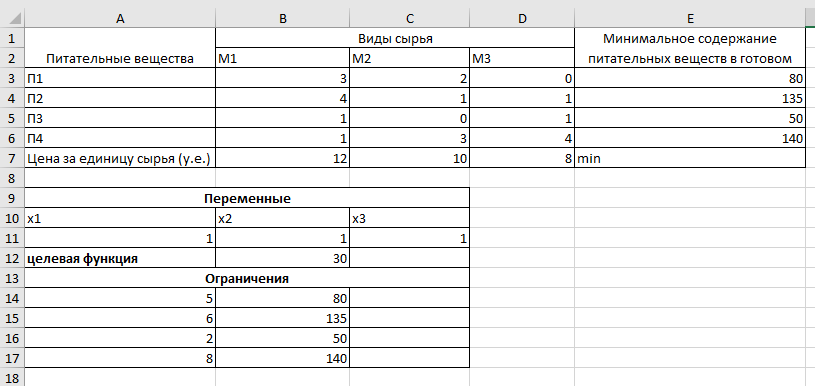


Рис. 20. Таблица для задачи 2

Аналогично прошлой задаче, используя команду “Поиск решения”, заполним открывшееся окно и нажмём кнопку “Найти решение” (см. рис. 21).

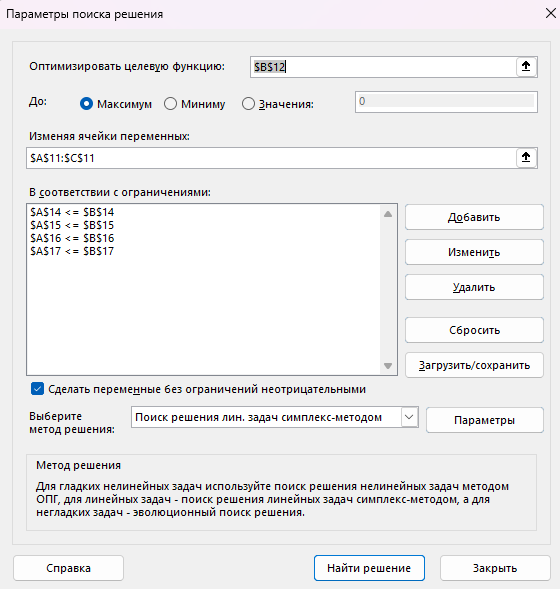


Рис. 21. Поиск решения для задачи 2

И получим обновлённую таблицу с результатом (см. рис. 22).

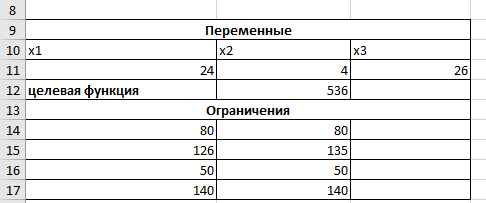


Рис. 22. Результаты для задачи 2

На основе полученных результатов можно сделать следующий вывод: минимальная по стоимости смесь сырья включает в себя пропорцию сырья М1:М2:М2 равную 24:4:26.

# Заключение

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)

В ходе выполнения учебной практики были изучены различные инструменты для анализа данных и решения оптимизационных задач. Особое внимание уделено программному пакету Orange и возможностям Excel для решения задач линейного программирования.

В целом, учебная практика позволила значительно расширить знания и навыки в области анализа данных и решения оптимизационных задач, используя современные информационные технологии и программное обеспечение.

# Список литературы

1. KNIME Analytics Platform [Электронный ресурс] : офиц. сайт. — URL: https://www.knime.com/ (дата обращения: 03.11.2024).
2. KNIME Hub Pricing [Электронный ресурс] : офиц. сайт. — URL: https://www.knime.com/knime-hub-pricing (дата обращения: 03.11.2024).
3. IBM SPSS [Электронный ресурс] : офиц. сайт. — URL: https://www.ibm.com/spss (дата обращения: 03.11.2024).
4. SAS: мы анализировали данные и обучали модели задолго до того, как это стало модным. // Блог компании SAS. — URL: https://habr.com/ru/companies/sas/articles/343140/ (дата обращения: 03.11.2024).
5. StatSoft Europe GmbH [Электронный ресурс] : офиц. сайт. — URL: https://www.statsoft.de/en/product/spotfire-statistica-ultimate-academic/ (дата обращения 03.11.2024).
6. ПРОГРАММА «STATISTICA»: методические указания к выполнению лабораторных работ / Н.И. Тебайкина. – Екатеринбург: ГОУ ВПО УГТУ–УПИ, 2006.
7. Minitab [Электронный ресурс] : офиц. сайт. — URL: https://www.minitab.com/en-us/ (дата обращения: 05.11.2024).
8. R Project [Электронный ресурс] : офиц. сайт. — URL: https://www.r-project.org/ (дата обращения: 05.11.2024).
9. Python [Электронный ресурс] : офиц. сайт. — URL: https://www.python.org/ (дата обращения: 06.11.2024).
10. Orange Data Mining [Электронный ресурс] : офиц. сайт. — URL: https://orangedatamining.com/ (дата обращения: 06.11.2024).
11. Loginom [Электронный ресурс] : офиц. сайт. — URL: https://loginom.ru/ (дата обращения: 06.11.2024).