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 Зернов Владислав Александрович группы 1ПИб-02-2оп-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.)

Содержание

[Введение 3](#_Toc153890784)

[Раздел 1. Работа в пакете Orange. Его возможности, результаты выполнения задания 11](#_Toc153890785)

[Раздел 2. Описание решения оптимизационной задачи в Excel 21](#_Toc153890786)

[Заключение 27](#_Toc153890787)

# Введение

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

Целью технологической практики на втором курсе является обзорное знакомство с некоторыми из этих пакетов. В дальнейшем часть из них будут использоваться как инструментарий.

**Изучение возможностей инструментов для анализа данных**

В данном разделе предоставлена информация о различных инструментах и пакетов, используемых при работе с данными.

**KNIME**

1. KNIME является бесплатным и открытым программным обеспечением. Для коммерческого использования также требуется лицензия, однако студентам и академическим учреждениям предоставляется бесплатный доступ к платным пакетам.
2. KNIME предназначен для автоматизации процессов обработки и анализа данных, а также для интеграции различных инструментов и технологий в рамках единой платформы.
3. С помощью KNIME можно выполнять математические расчеты, моделирование, статистические расчеты и многое другое.
4. KNIME предлагает широкий спектр инструментов для обработки данных, включая визуализацию данных, обработку текста, машинное обучение и т.д.
5. Для успешного применения KNIME необходимы знания в области анализа данных и машинного обучения, а также навыки работы с программным обеспечением.
6. Для работы с KNIME необходимо иметь базовые навыки программирования, так как программа работает на основе скриптов и алгоритмов.
7. KNIME работает с различными типами данных, такими как текстовые, числовые, временные ряды и другие. Программа также поддерживает работу с различными источниками данных, такими как базы данных, файлы и API.
8. Результатом работы KNIME может быть визуализированная информация о данных, модели машинного обучения, отчеты и другие результаты анализа

**SPSS**

1. SPSS является платным инструментом, однако для образовательных учреждений и студентов существуют бесплатные лицензии.
2. Инструмент предназначен для статистической обработки и анализа данных.
3. С помощью SPSS можно проводить различные виды расчетов, включая моделирование и статистический анализ.
4. SPSS предлагает набор инструментов для работы с данными, включая инструменты для визуализации и анализа.
5. Для использования SPSS необходимы знания в области статистического анализа и навыки работы с программами.
6. Для работы с SPSS требуется знание программирования, так как это статистический пакет программ.
7. SPSS работает с разными видами данных, включая числовые и текстовые, а также может обрабатывать данные из различных источников.
8. В результате работы с SPSS пользователь получает статистические отчеты и графики, отражающие результаты анализа данных.

**SAS**

1. SAS - это коммерческий инструмент для аналитики и обработки данных. Однако для студентов и учебных заведений предоставляются бесплатные версии с ограниченным функционалом.
2. Этот инструмент используется для анализа данных, статистического моделирования и прогнозирования.
3. Он предоставляет возможность проведения математических расчетов, моделирования и статистического анализа.
4. SAS предоставляет различные инструменты и функции для работы с данными, их визуализации, работы с текстами и базами данных.
5. Чтобы эффективно использовать SAS, требуются знания в статистике, анализе данных и опыт программирования на языке SAS.
6. Требуется умение программировать на языке SAS, так как этот инструмент основан на программировании.
7. Он работает с различными типами данных: числовыми, категориальными, текстовыми и другими, а также поддерживает различные источники данных, такие как базы данных, файловые системы и API-интерфейсы.
8. Результатами работы в SAS являются статистические отчеты, графики, таблицы, модели прогнозирования и другие аналитические материалы.

**STATISTICA 6.0**

1. STATISTICA является платным продуктом, но для студентов и преподавателей предлагаются скидки и специальные условия.
2. STATISTICA предназначена для статистического анализа данных, визуализации результатов, построения моделей и проведения прогнозов.
3. STATISTICA позволяет проводить различные статистические расчеты, включая линейные и нелинейные модели, анализ временных рядов и кластеризацию.
4. STATISTICA предлагает различные инструменты для анализа данных: описательная статистика, анализ зависимостей, регрессионный анализ, анализ выживаемости и т.д.
5. Для успешного использования STATISTICA необходимы знания в области статистики и математического анализа. Также требуются навыки работы с компьютером и знание программного обеспечения.
6. Для использования STATISTICA не требуется навыков программирования, но знание статистических методов и алгоритмов может быть полезным.
7. STATISTICA работает с различными форматами данных, включая .csv, .xls, .txt и другие. Она также может импортировать данные из различных баз данных и других источников.
8. В результате использования STATISTICA пользователь получает отчеты о проведенном анализе, содержащие статистические выводы и графики.

**Minitab**

1. Minitab является платным ПО, однако есть бесплатная студенческая версия с некоторыми ограничениями.
2. Minitab используется для выполнения статистического анализа данных.
3. Minitab позволяет выполнять различные виды статистического анализа: описательный, корреляционный, регрессионный, анализ дисперсии и т. д.
4. Minitab предлагает множество инструментов для выполнения статистического анализа данных. Например, можно выполнять описательную статистику, корреляционный анализ, регрессионный анализ и т. д.
5. Для успешного использования Minitab необходимы базовые знания статистики и умение работать с ПО.
6. Программирование не требуется для использования Minitab.
7. Minitab может работать с различными типами данных, включая числовые, текстовые, категориальные и т. д. Он также поддерживает работу с различными источниками данных, такими как базы данных, файлы и т. д.
8. Результатом работы в Minitab являются статистические отчеты с результатами анализа данных.

**R + R-Studio**

1. R и R-Studio являются бесплатными программами. R - это язык программирования для статистического анализа, а R-Studio - это среда для разработки и визуализации данных на R. Обе программы доступны для загрузки и использования бесплатно.
2. Эти инструменты предназначены для проведения статистического анализа данных, моделирования, обработки данных, создания визуализаций и отчетов.
3. R позволяет проводить широкий спектр математических вычислений, моделирования и статистических расчетов.
4. R-Studio предлагает инструменты для управления проектами, отладки кода, визуализации данных и работы с удаленными серверами.
5. Для эффективного использования R и R-Studio требуются знания в статистическом анализе, программировании на языке R, а также опыт работы с R-Studio.
6. Для эффективного использования этих инструментов необходимо умение программировать на языке R.
7. R и R-Studio работают с различными типами и форматами данных, включая текстовые файлы, базы данных, CSV-файлы и другие. Они также поддерживают работу с данными из различных источников, таких как интернет, API-интерфейсы и другие.
8. Результатом их работы являются статистические данные, графики, модели и другие аналитические результаты.

**Python**

1. Python является бесплатным программным обеспечением с открытым исходным кодом, доступным для загрузки и использования без каких-либо ограничений.
2. Python предназначен для широкого спектра задач, таких как веб-разработка, машинное обучение, анализ данных, автоматизация и многое другое.
3. Python предоставляет множество возможностей, включая проведение математических расчетов, создание моделей, выполнение статистического анализа и многое другое.
4. Python предлагает различные библиотеки и инструменты для работы с данными, такие как numpy, pandas, scipy, matplotlib и многие другие.
5. Чтобы успешно использовать Python, вам потребуются знания в области программирования, статистики, машинного обучения и других областей в зависимости от того, для чего вы планируете использовать этот язык.
6. Нужно уметь программировать на Python, чтобы использовать его эффективно.
7. Python может работать с разными типами данных и источниками, включая файлы, базы данных, API, веб-сервисы и многое другое.
8. В результате работы с Python вы можете получить разные виды результатов, такие как отчеты, графики, статистические данные и многое другое, в зависимости от задач, которые вы решаете.

**Orange**

1. Orange является бесплатным программным обеспечением и может быть загружен и использован без ограничений. Он также предлагает академическую лицензию для образовательных учреждений.
2. Orange предназначен для задач анализа данных, таких как визуализация, разведочный анализ данных, прогнозирование и т. д.
3. Возможности Orange включают широкий спектр инструментов для проведения математических расчетов, моделирования и статистических анализов.
4. Средства и инструменты, предлагаемые Orange, включают визуализацию данных, алгоритмы машинного обучения, обработку естественного языка и т.д.
5. Для успешного применения Orange требуются знания в областях статистики, анализа данных, машинного обучения.
6. Умение программировать не требуется для использования Orange. Весь процесс анализа данных осуществляется через графический интерфейс пользователя.
7. Orange может работать с различными источниками данных, такими как CSV-файлы, базы данных (SQL и NoSQL), веб-сервисы, API и т. д.
8. Результат работы в Orange включает визуализацию данных, прогнозы, модели машинного обучения и т.д.

**Loginom**

1. Loginom является платным программным обеспечением, но есть бесплатная версия для некоммерческого использования.
2. Программа предназначена для автоматизации процессов обработки данных, анализа и построения моделей.
3. Включает в себя возможности проведения математических расчетов, статистического анализа, моделирования.
4. Loginom предлагает различные средства и инструменты для работы с данными. Например: визуальное программирование, интеграция с другими инструментами и источниками данных.
5. Для успешного использования программы необходимы знания в области статистики, анализа данных и машинного обучения.
6. Программирование требуется для создания и настройки собственных моделей,но не более.
7. Loginom может работать с различными типами данных, включая текстовые, числовые, графические. Источниками могут служить файлы, базы данных, веб-сайты.
8. В результате работы программы получаются модели машинного обучения, результаты анализа данных, визуализации.

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

Создаем датасет и выбираем Housing. Для визуализации используем Data Table(рис.1).

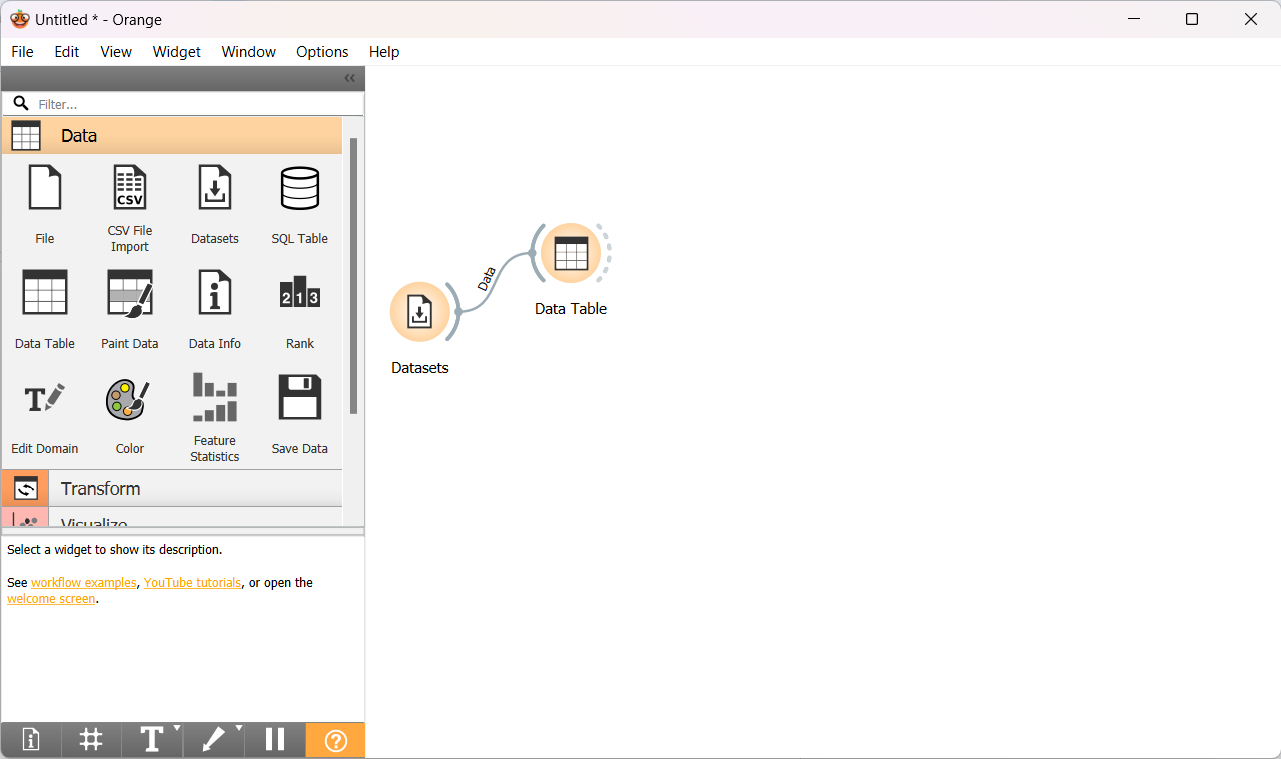
****

Рис. 1. Создание датасета

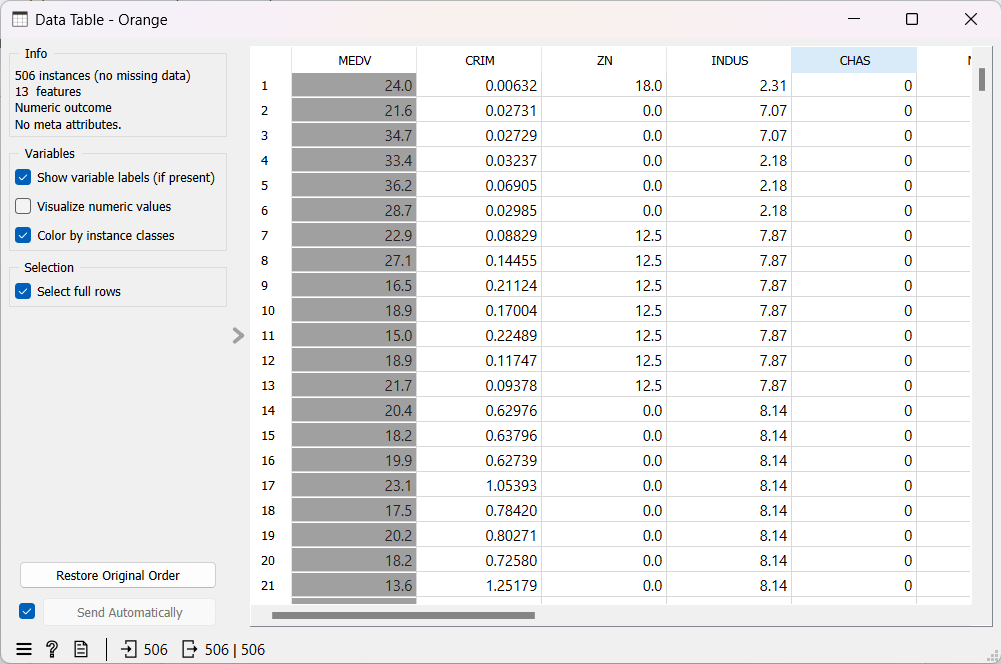
Просмотреть данные можно через виджет data table(рис.2) ****

Рис.2. Просмотр данных

Добавил Distribution для визуализации(рис.3, 4)

****

Рис. 3. Добавления виджета Distribution

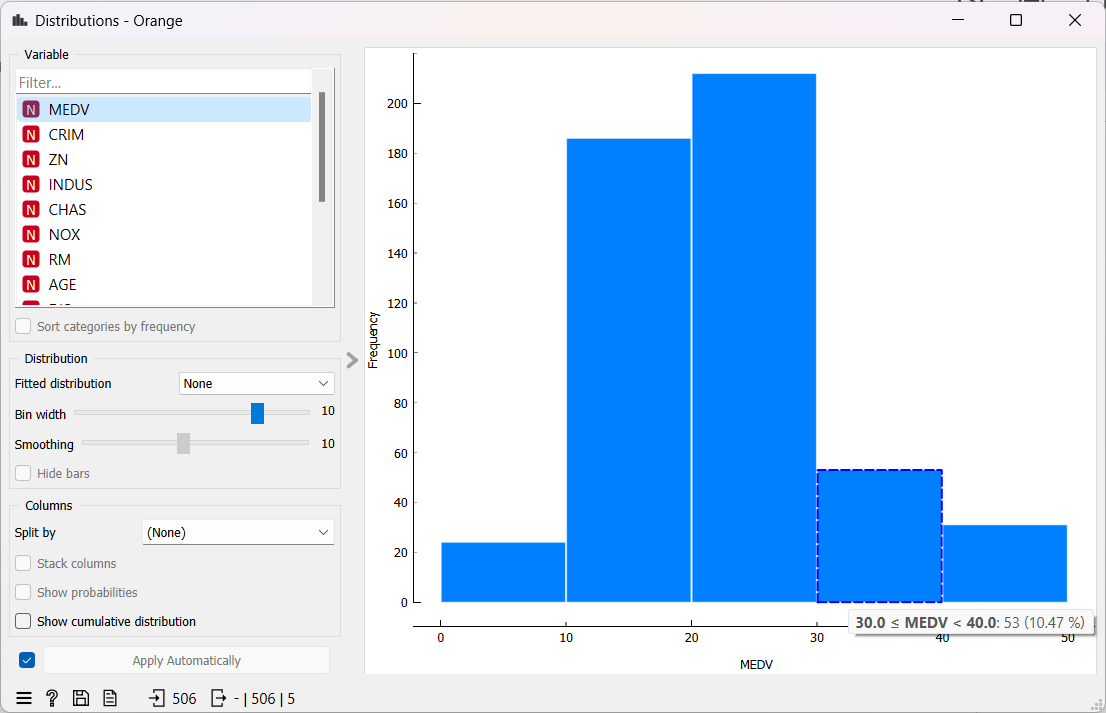
****

Рис.4 Просмотр визуализации

Для того, чтобы построить регрессию, нужно добавить виджеты Linear Regression и Test and Score (Рис. 5). Чтобы просмотреть результат, необходимо нажать на виджет Test and Score (Рис. 6).

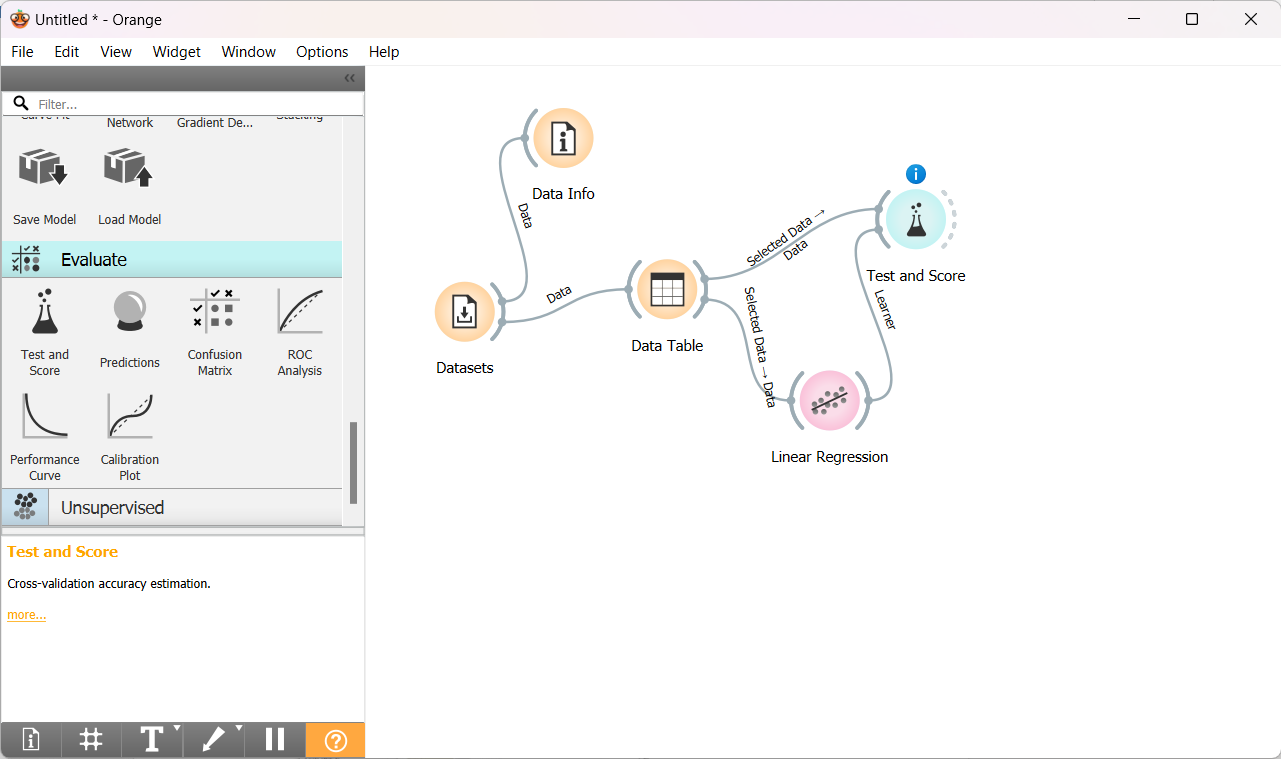
****

Рис. 5. Добавление новых виджетов

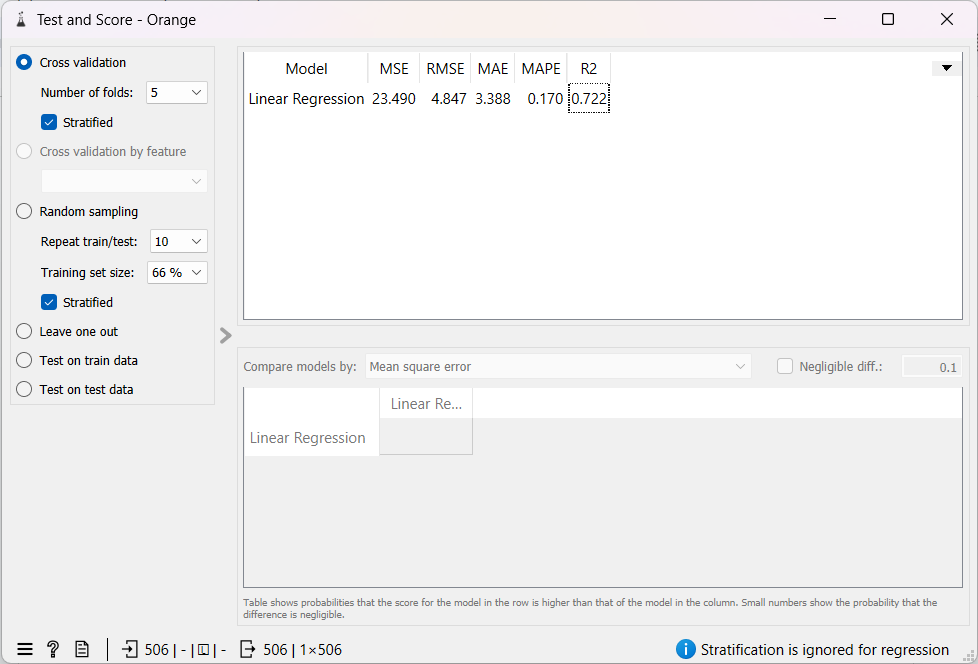
****

Рис. 6. Результат построения регрессии

Добавление виджета Preprocess и еще одного Data Table для улучшения расчетов. При нажатии на новый виджет Data Table откроется специальное меню, где можно проанализировать коэффициенты регрессии и то, как они влияют на результат (Рис. 8, 9).

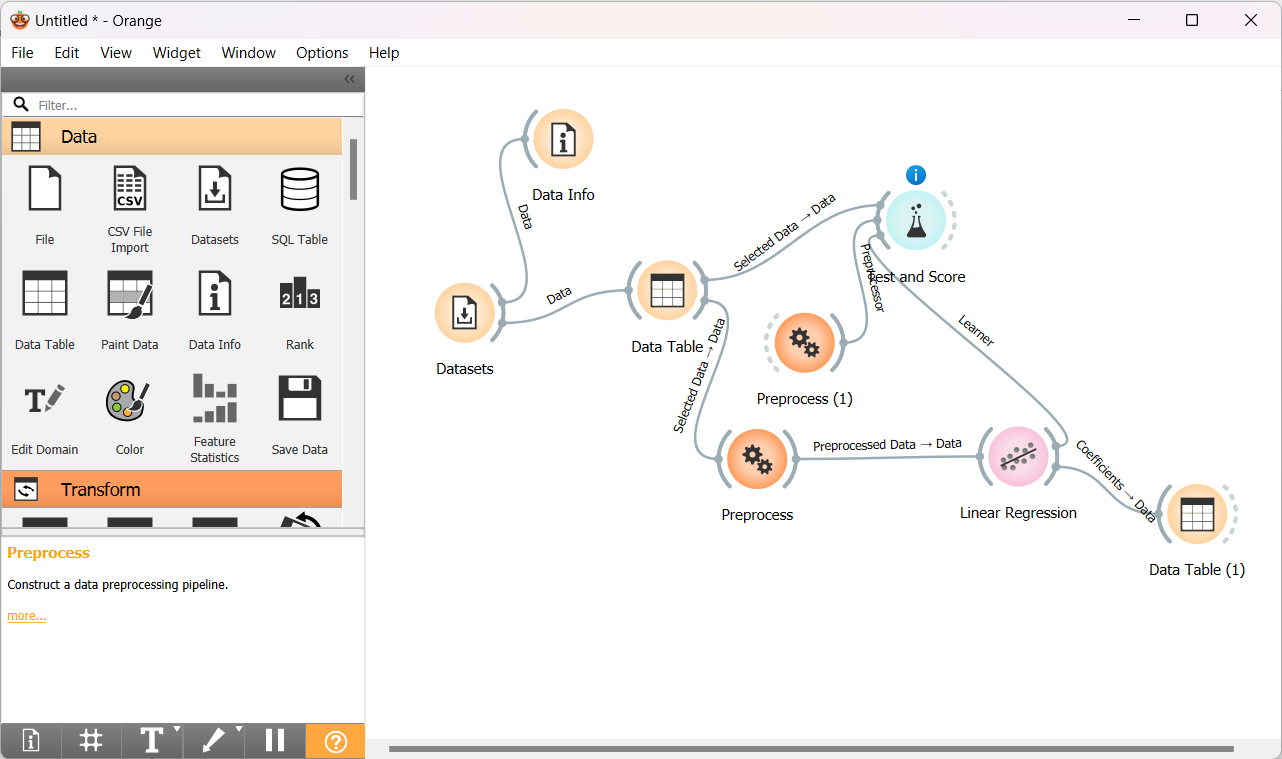


Рис. 8. Добавление виджетов

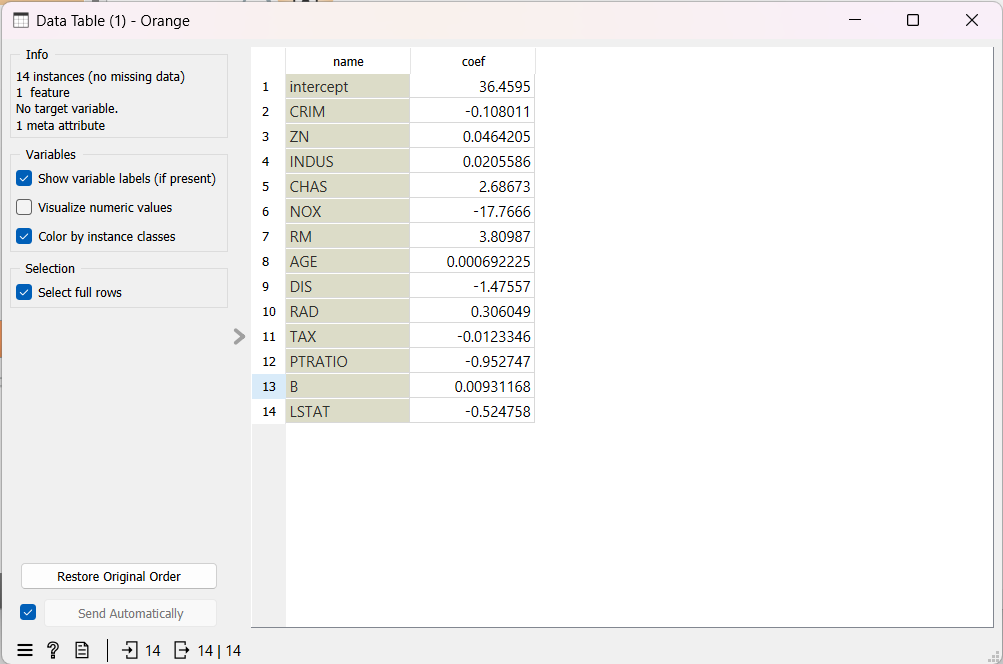


Рис. 9. Результат препроцессинга

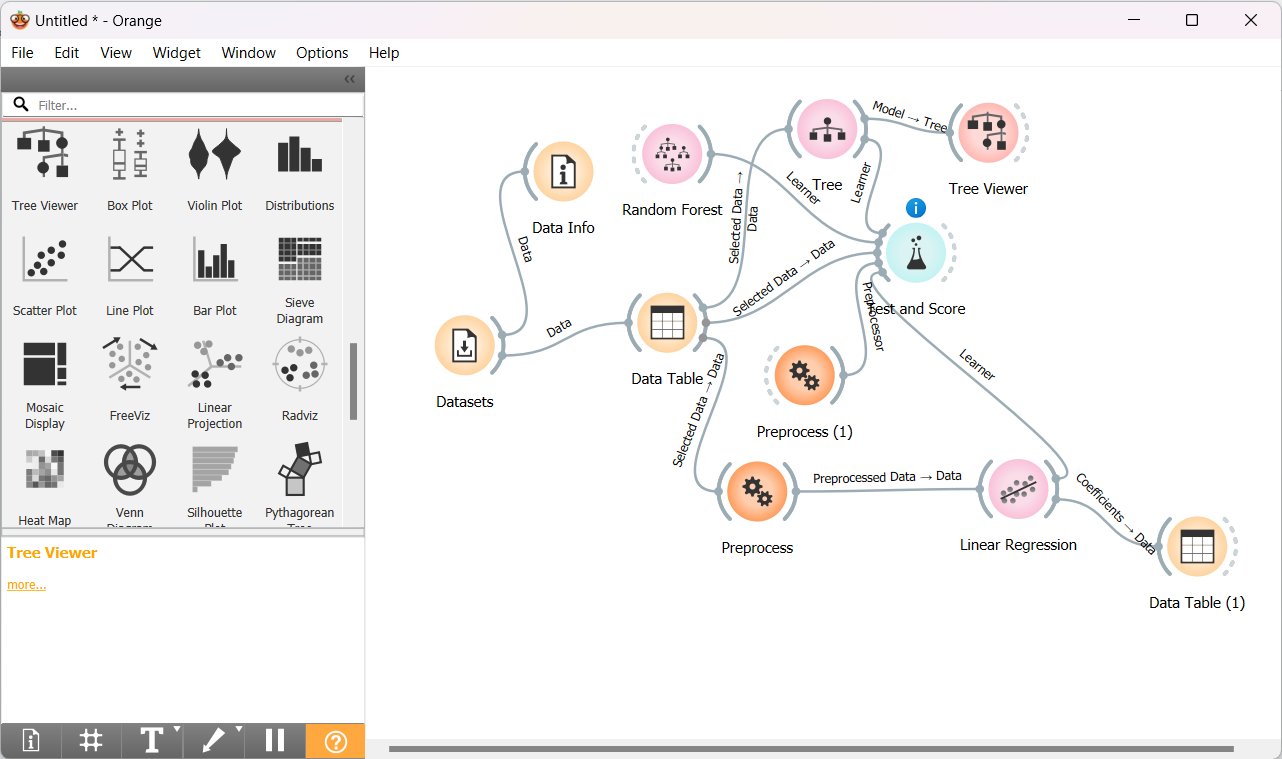
Данные можно визуализировать в виде дерева. Для этого нужно добавить виджеты Random Forest, Tree и Tree Viewer (Рис. 10). При нажатии на Tree Viewer данные будут визуализированы в соответствующем виде (Рис. 11). 

Рис. 10. Добавление виджетов

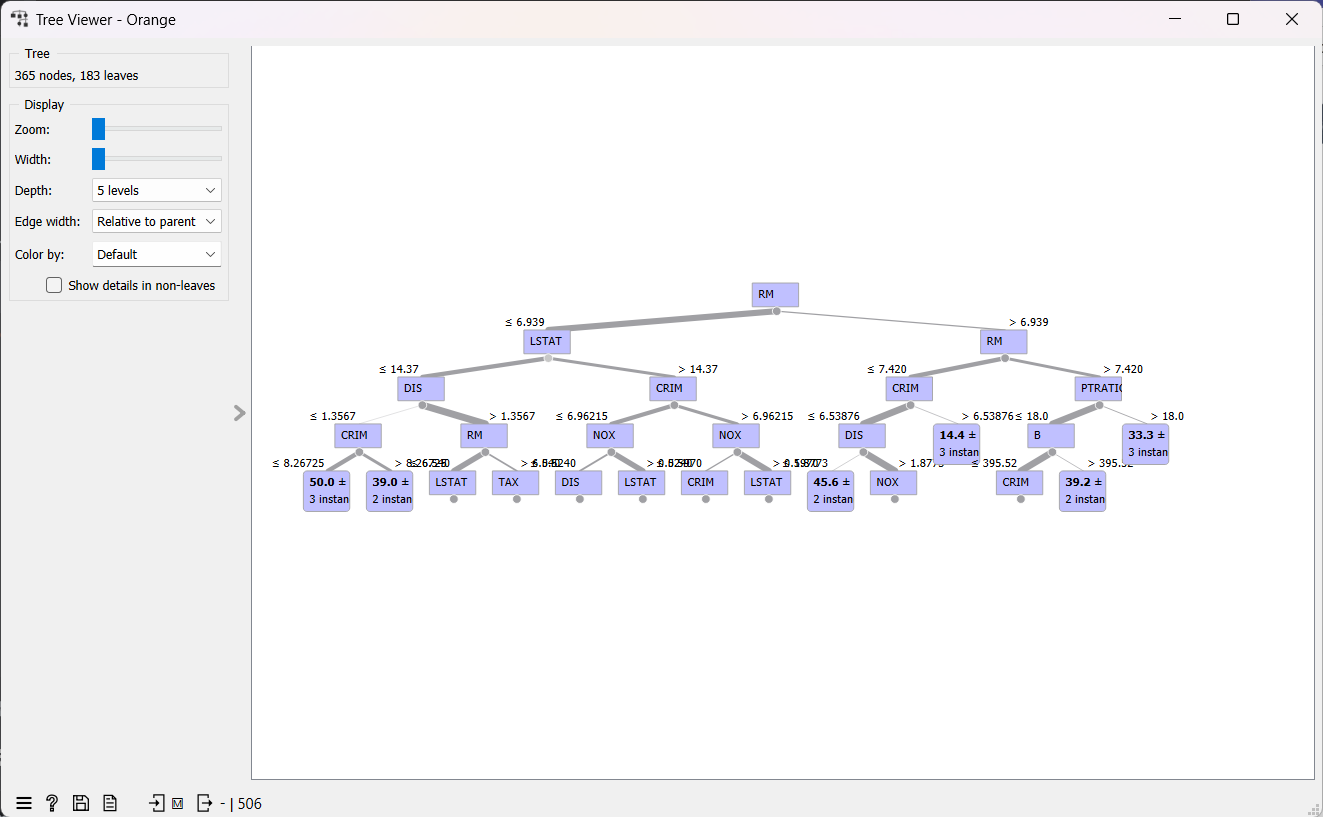


Рис. 11. Визуализация данных в виде дерева

**Заключение по работе с Orange**

Данное ПО обладает удобным инструментарием и приятным интерфейсом. Обладает визуализацией данных. Для студентов данное ПО будет помогать в работе с какими-то статистическими заданиями.

**Заключение по работе с инструментами для анализа данных**

**KNIME**

Программное обеспечение KNIME является удобным инструментом для анализа данных с широким спектром инструментов для обработки и визуализации данных. Для его использования требуется определенная подготовка, но благодаря интуитивно понятному интерфейсу освоить программу может пользователь с разным уровнем подготовки. Инструменты анализа данных позволяют решать различные задачи и могут быть адаптированы под конкретные потребности. Таким образом, KNIME является отличным выбором для анализа данных и машинного обучения.

**SPSS**

SPSS - удобный и эффективный инструмент для анализа данных с большим количеством статистических функций. Требует определенной подготовки пользователя, но имеет интуитивно понятный интерфейс, но у KNIME и Orange он приятнее. Позволяет решать различные статистические задачи. Подходит для анализа данных, научных исследований и бизнеса.

**SAS**

SAS является мощным и удобным инструментом для работы с данными с широким спектром функций и методов анализа. Программа имеет интуитивно понятный интерфейс, но требует некоторой подготовки для использования. Аналитические инструменты SAS позволяют решать различные задачи - от простой обработки данных до сложных моделей машинного обучения. Они могут быть адаптированы под нужды пользователя и гарантируют высокую точность результатов. В целом, SAS отлично подходит для аналитиков данных, исследователей и специалистов в области машинного обучения.

**STATISTICA 6.0**

STATISTICA 6.0 представляет собой мощное и надежное программное обеспечение для статистического анализа данных. Этот инструмент предлагает широкий спектр функций и инструментов, которые позволяют пользователям обрабатывать и анализировать данные. Некоторые из ключевых особенностей STATISTICA 6.0 включают описательную статистику, корреляционный анализ, регрессионный анализ, анализ выживаемости и многое другое. Кроме того, программа предоставляет интуитивно понятный интерфейс, что делает ее доступной для пользователей с различным уровнем знаний и опыта. В целом, STATISTICA 6.0 подходит для специалистов, которые занимаются статистическим анализом данных, проводят научные исследования и занимаются бизнес-анализом.

**Minitab**

Minitab - это удобный и эффективный инструмент для анализа данных со множеством функций и статистических анализов. Для работы с программой требуется определенная статистическая подготовка, однако Minitab имеет простой и интуитивный интерфейс с обучающими материалами для облегчения процесса обучения. Minitab идеально подходит для статистического анализа данных в научных исследованиях, образовании и бизнесе.

**R + R-Studio**

R и R-Studio представляют собой удобные и эффективные инструменты для анализа данных и статистической обработки. Они предлагают широкий спектр функций для работы с различными типами данных, включая линейную регрессию, кластерный анализ, анализ временных рядов и многое другое. Для использования R и R-Studio требуется определенная статистическая и программистская подготовка, однако благодаря их простоте и интуитивности, эти инструменты доступны для пользователей с разными уровнями навыков. R и R-Studio идеально подходят для статистической обработки данных в научных исследованиях, анализе бизнеса и образовательных целях.

**Python**

Python является удобным и популярным языком программирования для анализа данных и создания статистических моделей. Он имеет широкий спектр библиотек и инструментов, таких как pandas, numpy, scipy, matplotlib, которые упрощают процесс обработки и визуализации данных. Для использования Python не требуется специальная подготовка, так как он имеет простой и понятный синтаксис. Python подходит для анализа данных в науке, бизнесе и образовании.

**Loginom**

Loginom представляет собой удобный и функциональный инструмент для работы с бизнес-процессами и анализом данных. Он обладает широким спектром инструментов для автоматизации процессов, анализа данных и построения моделей машинного обучения. Для работы с Loginom требуется определенный уровень подготовки в области бизнес-анализа и статистики. Loginom подходит для анализа данных в различных отраслях, включая финансы, маркетинг и производство.

# Раздел 2. Описание решения оптимизационной задачи в Excel

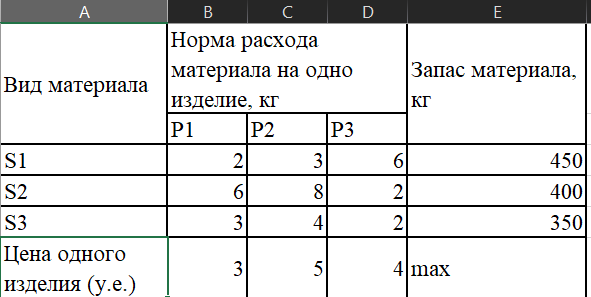
Задача 1:

Для изготовления трех 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 1. Составить план выпуска изделий, обеспечивающих их максимальный выпуск по стоимости.

Таблица 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Вид материала | Норма расхода материала на одно изделие, кг | | | Запас материала, кг |
| P1 | P2 | P3 |
| S1 | a11 | a12 | a13 | b1 |
| S2 | a21 | a22 | a23 | b2 |
| S3 | a31 | a32 | a33 | b3 |
| Цена одного изделия (у.е.) | c1 | c2 | c3 | max |

Вариант 10



1. Составить математическую модель задачи:

Введем неизвестные:

x1- кол-во продукции P1

x2- кол-во продукции P2

x3- кол-во продукции P3

Запишем ограничения задачи:

Зпишем целевую функцию:

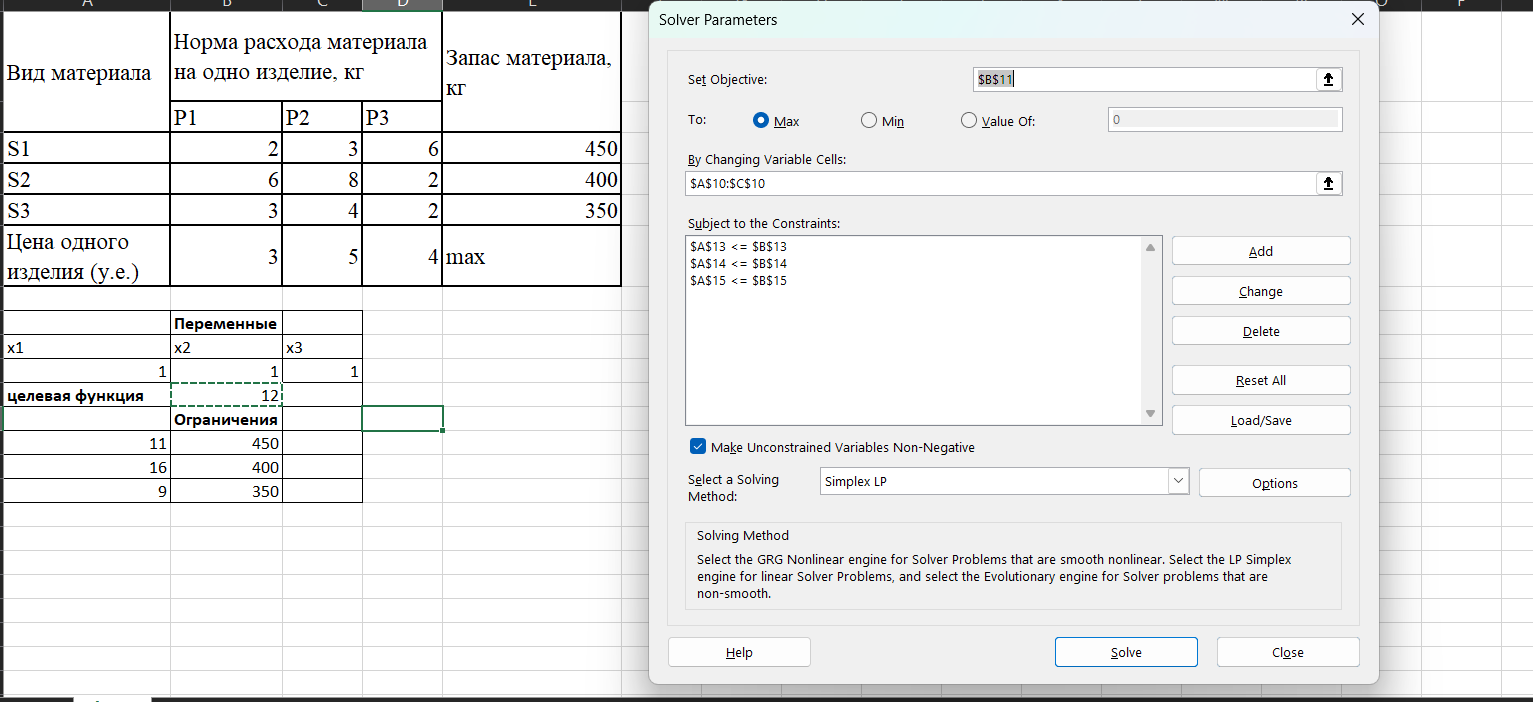


Рис. 12. Поиск решения

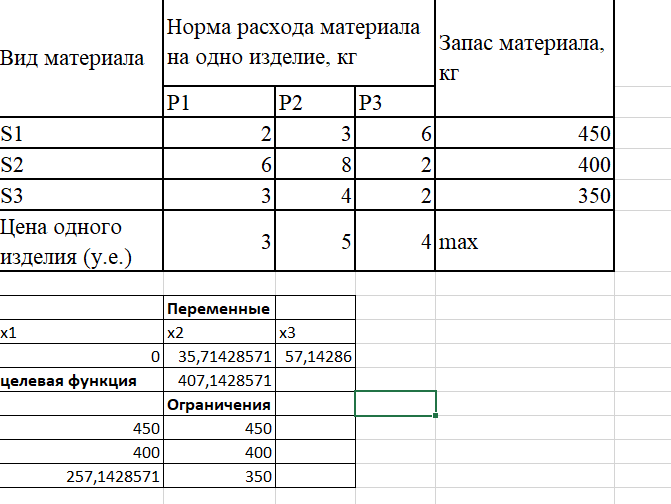


Рис. 13.Результат решения

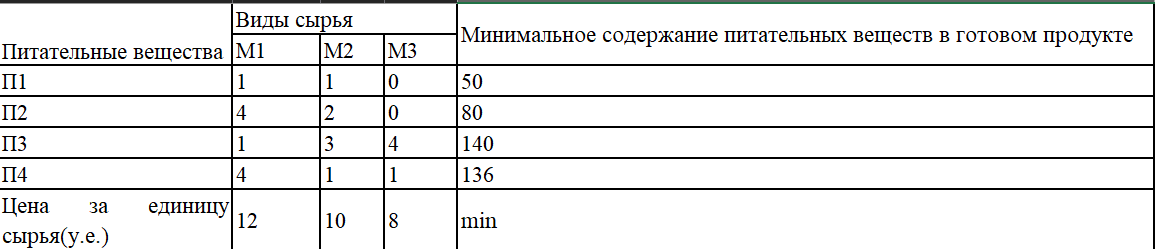
На основе полученных результатов можно сделать вывод: предприятию надо выпустить продукции вида P2 в количестве 35 усл. Единиц, вида P3 надо выпустить в количестве 57, продукцию вида P1 выпускать невыгодно. При таком плане выпуска продукции с учетом имеющихся ограничений будет получена максимальная прибыль в размере 407 усл. единиц.

Задача 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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Питательные вещества | Виды сырья | | | Минимальное содержание питательных веществ в готовом продукте |
| М1 | М2 | М3 |
| П1 | a11 | a12 | a13 | b1 |
| П2 | a21 | a22 | a23 | b2 |
| П3 | a31 | a32 | a33 | b3 |
| П4 | а41 | а42 | а43 | b4 |
| Цена за единицу сырья(у.е.) | c1 | c2 | c3 | min |

Вариант 10:



Запишем ограничения задачи:

Зпишем целевую функцию:

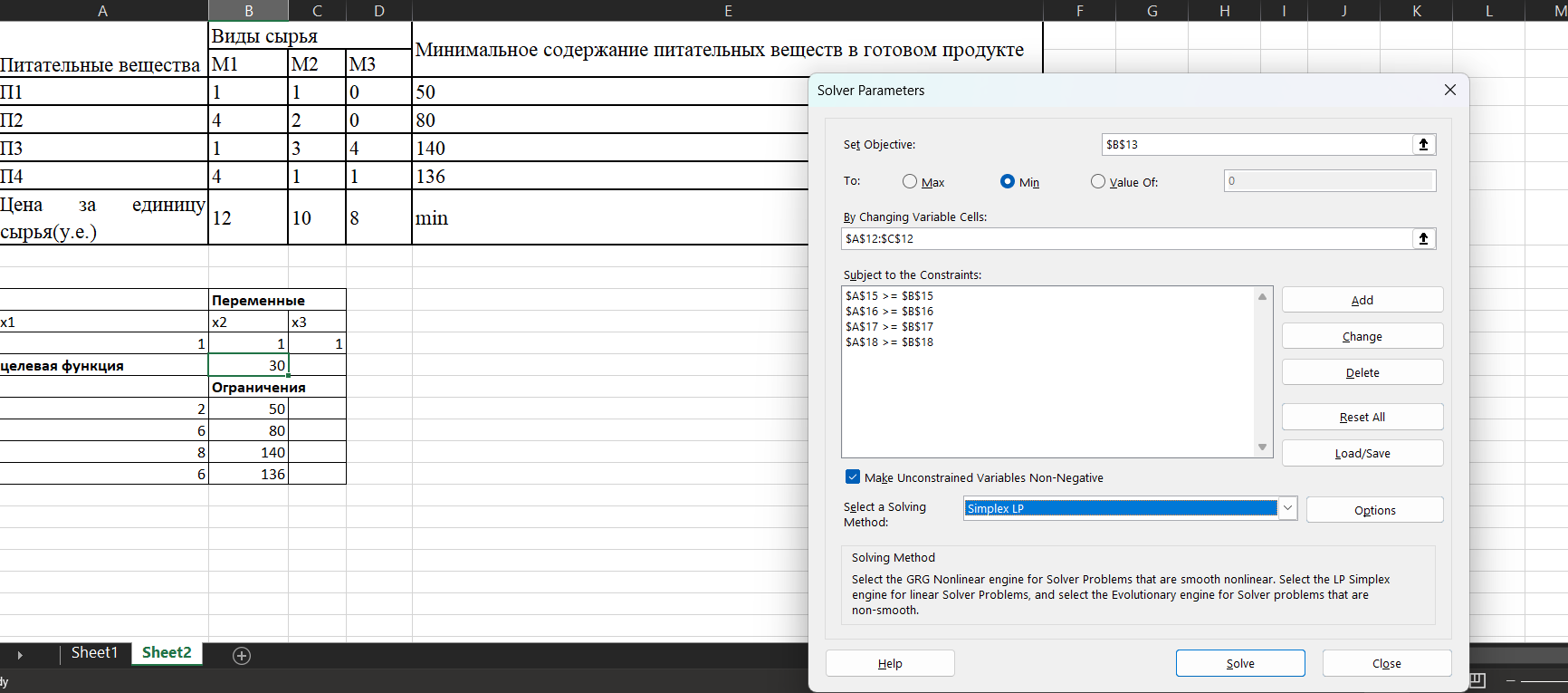


Рис. 14. Поиск решения

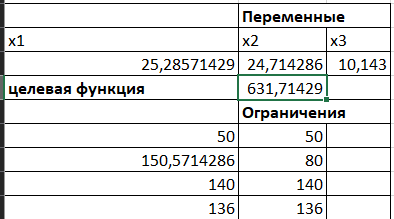


Рис. 15. Результат решения

На основе полученных результатов можно сделать следующий вывод: необходимо выпустить сырье первого вида M1 в количестве 26 усл. единицы, второго вида M2 в количестве 25 у. е., третьего вида M3 в количестве 11 у. е. При таком плане выпуска пищевых концентратов с учетом имеющихся ограничений будет получена минимальная цена смеси в размере 632 у. Е

# Заключение

В процессе прохождения технологической практики на кафедре МПО ЭВМ ЧГУ получилось ознакомиться с пакетами анализа данных Orange, STATISTICA, SPSS, KNIME, Minitlab, Loginom, R+R-Studio и SAS. Также удалось поработать с пакетом Orange и протестровать его возможности.

Также по ходу решения оптимизационных задач были получены и закреплены навыки работы с пакетом анализа Excel.

В процессе выполнения заданий по технологической 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).