ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ МОСКОВСКОЙ ОБЛАСТИ

«Университет «Дубна»

ИНСТИТУТ СИСТЕМНОГО АНАЛИЗА И УПРАВЛЕНИЯ

Кафедра системного анализа и управления

КУРСОВАЯ РАБОТА

по дисциплине

«Теория принятия решений»

Принятие решений в задаче распознавания образов с использованием алгоритма *k*-ближайших соседей

ТЕМА: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

(наименование темы)

2013

Выполнил: студент группы \_\_\_\_\_\_\_\_\_

Ржевский В. С.

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

(Ф.И.О.)

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

(подпись студента)

Руководитель:

Доцент Кирпичева Е. Ю.

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

(ученая степень, ученое звание, занимаемая должность, ФИО)

Дата: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Оценка: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

(подпись руководителя)

Дубна, 2021

**Содержание**

[Введение 3](#_Toc72806975)

[Постановка задачи 4](#_Toc72806976)

[Теоретическая часть 5](#_Toc72806977)

[Практическая часть 7](#_Toc72806978)

[*Описание интерфейса* 7](#_Toc72806979)

[*Распознавание образов* 7](#_Toc72806980)

[Заключение 9](#_Toc72806981)

[Список литературы 10](#_Toc72806982)

[Приложение 11](#_Toc72806983)

# Введение

Теория принятия решений — область исследования, вовлекающая понятия и методы математики, статистики, экономики, менеджмента и психологии, которая изучает закономерности выбора людьми путей решения разного рода задач, а также исследует способы поиска наиболее выгодных из возможных решений.

Данная курсовая работа посвящена распознаванию образов методом *k*-ближайших соседей. Для реализации данного метода была разработана программа в среде разработки *Microsoft Visual Studio* 2019 на языке *C#* с использованием интерфейса *Windows Forms*.

# Постановка задачи

**Цель работы:** разработать программу, реализующую алгоритм *k*-ближайших соседей, для решения практических задач.

**Исходные данные:** информация, собранная в процессе выполнения курсовой работы, теоретические знания об алгоритме *k*-ближайших соседей, а также среда программирования *Microsoft Visual Studio* 2019.

**Априорные модельные представления:** программа позволяет пользователю задать данные, имена критериев, способ нахождения расстояния, количество соседей.

**Результат курсовой работы:** программа будет определять принадлежности объектов к одному из образов.

# Теоретическая часть

В данной работе используются следующие понятия:

Образ — совокупность объектов, объединенная некоторыми общими свойствами.

Объект — точка в многомерном признаковом пространстве. Объектами могут быть любые сущности, которые необходимо подвергнуть классификации.

Материал обучения (МО) — объекты, на которых измерены прямые свойства.

Материал экзамена (МЭ) — объекты, на которых не измерены прямые свойства.

Свойство (признак) — некоторая характеристика объекта.

Таблица объект-свойство (ТОС) — таблица, содержащая информацию о свойствах отдельных объектов, принадлежащих одному классу.

Метрика — любая функция , определяющая расстояние между двумя объектами и удовлетворяющая следующим условиям:

1. *.*

Метод ближайшего соседа является самым простым алгоритмом классификации. Классифицируемый объект относится к тому образу, которому принадлежит ближайший объект материала обучения.

**Метод *k*-ближайших соседей.** Для повышения надежности классификации объект относят к тому классу, которому принадлежит большинство из его соседей. В задачах с двумя образами число соседей берут нечетным, чтобы не возникало ситуаций неоднозначности, когда одинаковое число соседей принадлежат разным образам.

Описание алгоритма:

1. Пусть задана обучающая выборка пар "объект-ответ"
2. На множестве объектов задана метрика . Примеры метрик:

* Обобщенная метрика Минковского
* Евклидова метрика
* Расстояние Манхэттена

.

1. Для произвольного объекта материала экзамена *u* расположим объекты материала обучения в порядке возрастания расстояний до *u*.

, где через обозначается тот материал обучения, который является *i*-м соседом объекта *u*.

1. Аналогичное обозначение введем и для ответа на *i*-м соседе: . Таким образом, произвольные объект *u* порождает свою перенумерацию выборки.
2. Чтобы сгладить шумовое влияние выбросов, будем классифицировать объекты путем голосования по *k* ближайшим соседям. Каждый из соседей голосует за отнесение объекта *u* к своему классу .
3. Алгоритм относит объект *u* к тому классу, который наберет большее число голосов:

.

**Практическая часть**

## *Описание интерфейса*

При запуске программы пользователь видит перед собой форму (см. Приложение, рис. 1), в левой части которой расположены кнопки выбора способа решения, выбора весов, имена признаков, выбор количества соседей и кнопка решения.

Далее будут приведены части кода, описывающие работу программы. При запуске программы пользователь выбирает способ ввода значений: графический (см. Приложение, рис. 2) или через таблицу, также пользователь может выбрать случайный ввод значений (см. Приложение, рис. 3). После ввода значений, пользователь может выбрать имена признакам в соответствующих полях, по умолчанию признаки имеют имена «1й признак» и «2й признак», количество соседей, изначально равное 5, выбор метрики: Минковского, Евклидова или Манхэттена, которая передает в функцию *Metrick\_calculate* индекс выбранной метрики (см. Приложение, рис. 4).

Затем, после ввода всех необходимых значений и нажатия на кнопку «Определить», происходит проверка на корректный ввод всех объектов, материалов экзамена, весов и количества соседей, если все значения введены верно, то создается массив *distances*, в который записывается расстояние от материала экзамена до каждого материала объекта с помощью выбранной метрики. После заполнения массива происходит его сортировка и заполнение, для каждого материала экзамена происходит ввод в список *sequences*. Далее, создаются массивы *tmp\_1* и *tmp\_2*, размером равным количеству выбранных соседей, и задаются, какому объекту соответствует *i-*й по счету сосед в списке *sequences*. Затем, создаются две переменные *num\_1* и *num\_2*, которые суммируют количество соседей, перемноженный на веса, для каждого материала обучения. Все посчитанные значения: *num\_1*, *num\_2*, *tmp\_1,* *tmp\_2* и номер материала экзамена добавляются в список result\_list. Наконец, в поле *textBox\_result* выводится результат, если число *num\_1* больше числа *num\_2*, то данный материал экзамена принадлежит первому образу и наоборот (см. Приложение, рис. 5).

## *Распознавание образов*

Для демонстрации работы, решим задачу классификации пригодных для обитания планет. Выберем табличный ввод значений, вес оставим равным 0.5, зададим имена признакам: «Эффективная земная орбита», «масса (относительно Земли)», количество соседей установим равное 3, метрика — Евклидова. Первое свойство показывает, насколько конкретная планета расположена в обитаемой зоне, измеряется в астрономических единицах, Земля имеет 1 а. е. Обитаемая зона — диапазон расстояний от звезды, в пределах которого температура на поверхности планеты лежит в пределах от 0 до 100° С. При нормальном давлении атмосферы это открывает возможность существованию жидкой воды, а значит, и жизни в ее нынешнем понимании [4]. Второе свойство отображает массу планеты относительно Земли, чем больше масса, тем выше давление и ниже пригодность для обитания, а чем меньше, тем ниже притяжение, от которого зависит удержание атмосферы. В первый материал обучения добавим экзопланеты в обитаемой зоне или близкой к ней, во второй материал добавим данные из списка экзопланет, открытых телескопом *Kepler* [3, 5]. В материал экзамена введем значения для трех планет (см. Приложение, рис. 6). После подсчета всех значений и вывода на экран, видим, что первые две неизвестные планеты относятся к группе пригодной для жизни, а третья — нет. Также можно нажать на кнопку «График», где будет изображено распределение планет по группам (см. Приложение, рис. 7).

# Заключение

В ходе выполнения данной курсовой работы был реализован алгоритм *k-*ближайших соседей для распознавания образов. Проведено знакомство с теорией принятия решения в задачах распознавания образов. Разработана программа для решения задач и, с помощью данной программы, разобран конкретный пример распознавания образов с помощью разработанной программы.

# Список литературы

1. Hefin I. Rhys Machine Learning with R, the tidyverse, and mlr — Shelter Island: Manning, 2020. — С. 59-65.
2. Документация по *Windows forms* для .*NET 5* [Электронный ресурс] / Сайт компании *Microsoft*. Электрон. Текстовые дан.; Режим доступа: https://docs.microsoft.com/ru-ru/dotnet/desktop/winforms/?view=netdesktop-5.0, свободный.
3. Каталог экзопланет [Электронный ресурс] / Сайт компании *exoplanet TEAM*. Электрон. Текстовые дан.; Режим доступа: http://exoplanet.eu/, свободный.
4. Научно-популярный проект «Элементы» [Электронный ресурс] / Статья Владимира Сурдина «Зона жизни». Электрон. Текстовые дан.; Режим доступа: https://elementy.ru/nauchno-populyarnaya\_biblioteka/432211, свободный.
5. Список миссий телескопа *Kepler* [Электронный ресурс] / Сайт космического агентства *NASA*. Электрон. Текстовые дан.; Режим доступа: https://kepler.nasa.gov/Mission/discoveries/, свободный.

# Приложение

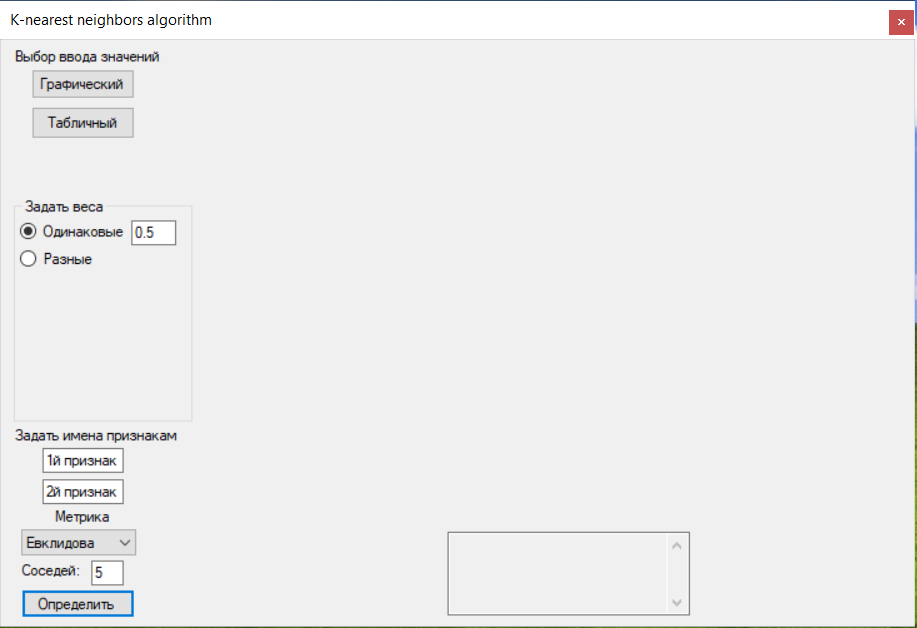


Рис. 1. Стартовое окно

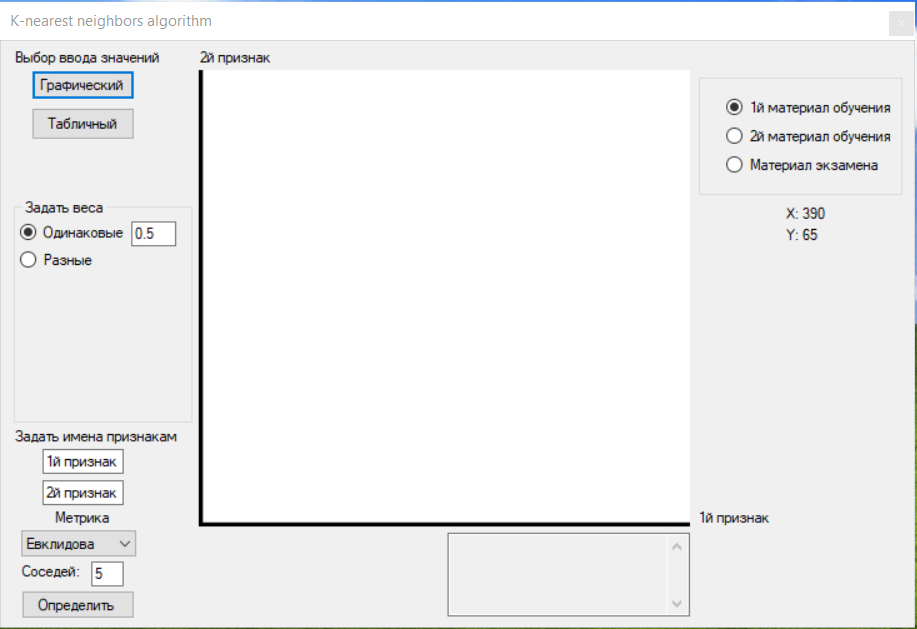


Рис. 2. Графический способ ввода

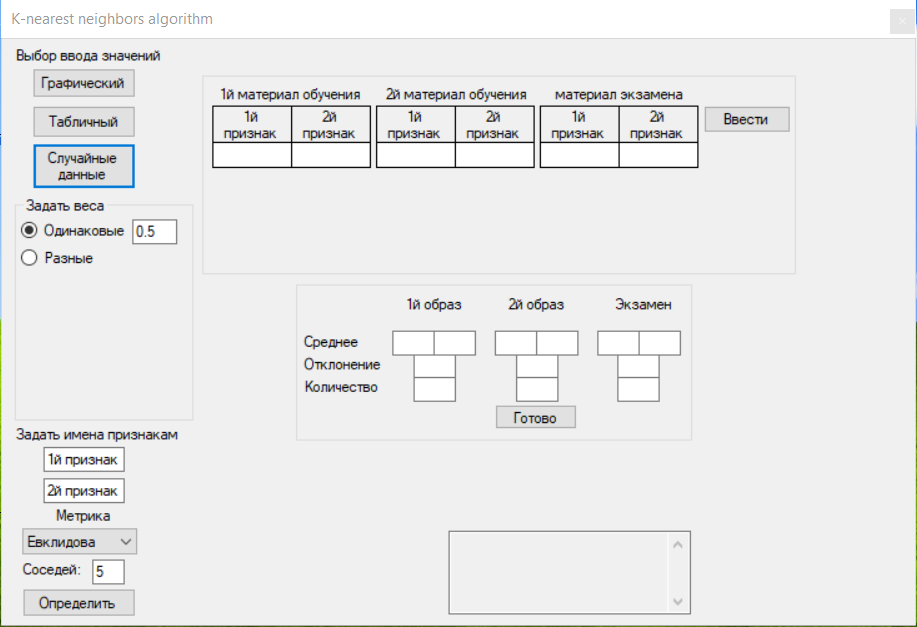


Рис. 3. Табличный способ ввода с формой для случайного ввода значений

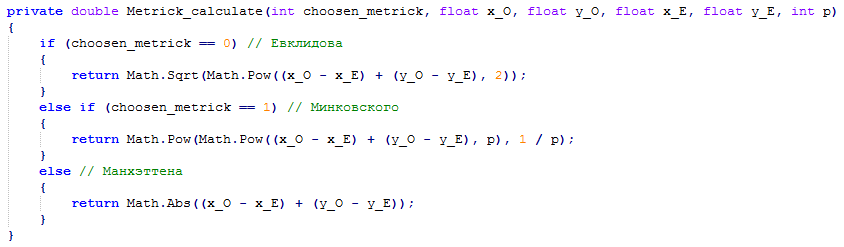


Рис. 4. Функция выбора метрики

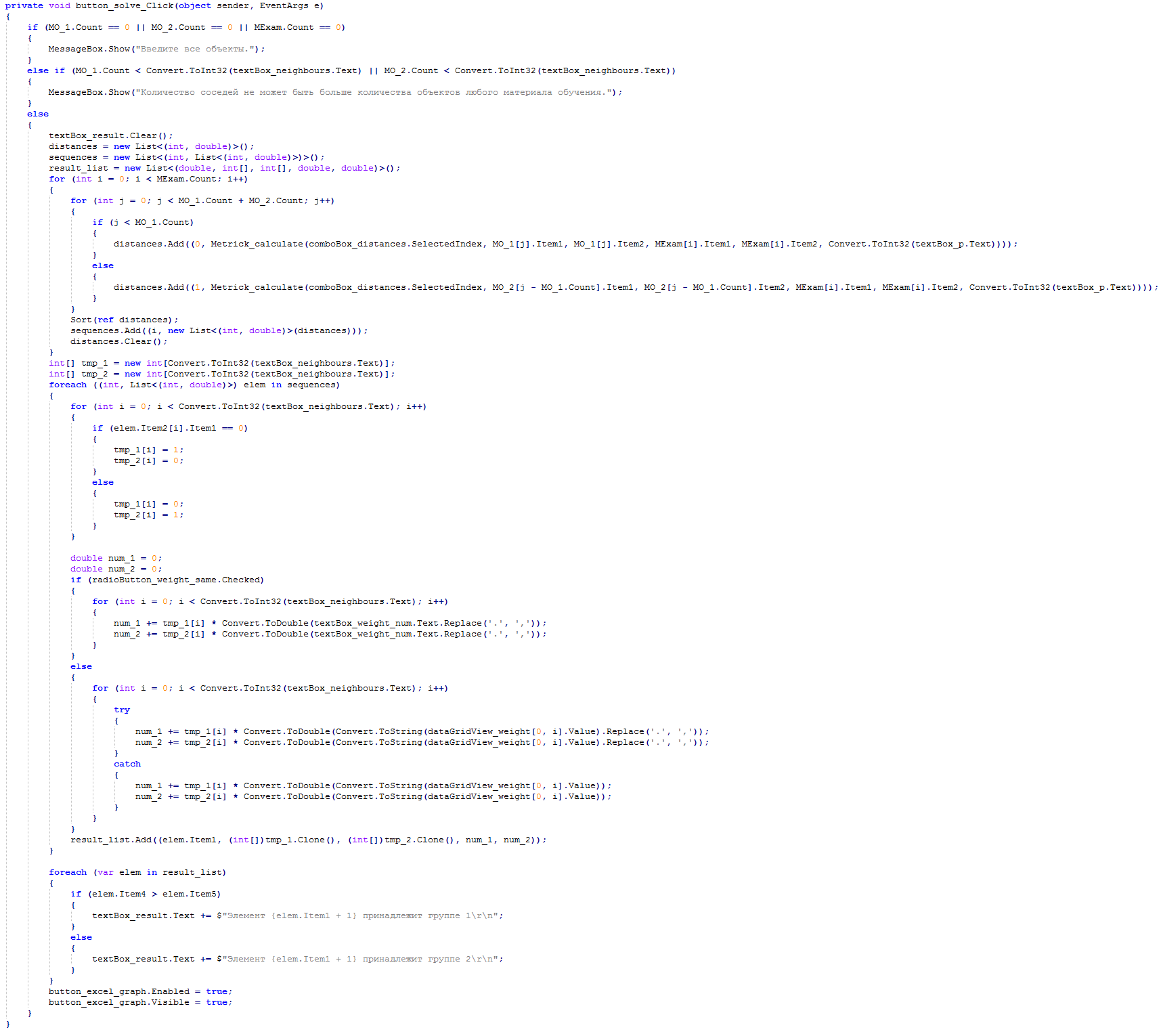


Рис. 5. Функция решения

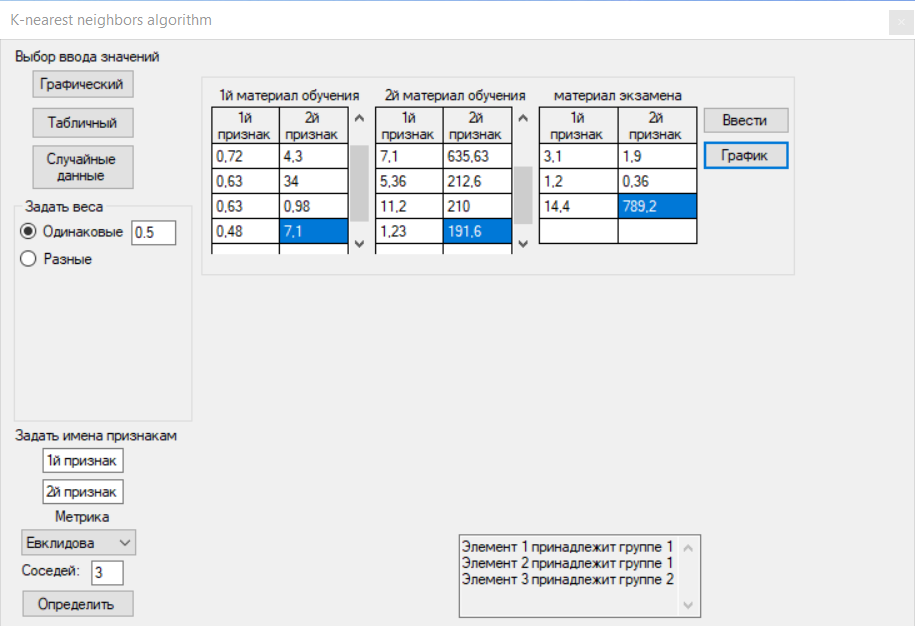


Рис. 6. Ввод и обработка данных

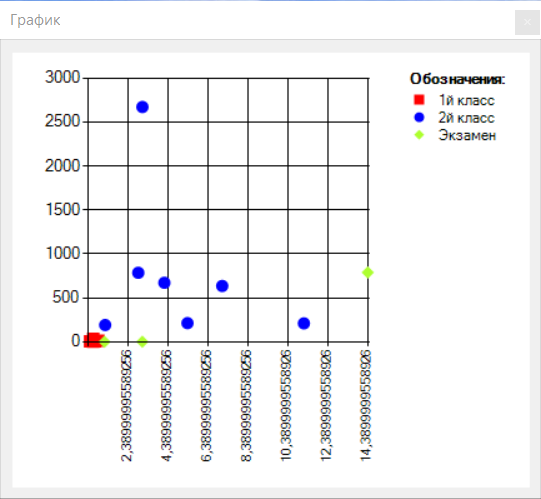


Рис. 7. Итоговый график