|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА – Российский технологический университет»** |
| **РТУ МИРЭА** |
| Институт кибербезопасности и цифровых технологий |
| *(наименование института, филиала)* |
|  |
| Кафедра информатики |
| *(наименование кафедры)* |

**КУРСОВАЯ РАБОТА**

| по дисциплине | **Основы машинного обучения** |
| --- | --- |
|  | *(наименование дисциплины)* |

| **Тема курсовой работы** | **Реализация алгоритмов решающих деревьев** |
| --- | --- |
|  | |

| **Студент группы** | БФБО-05-22 Нахамкин Константин Дмитриевич |  |  |
| --- | --- | --- | --- |
|  | *(учебная группа, фамилия, имя, отчество студента)* |  | *(подпись студента)* |

| **Руководитель курсовой работы** | зав. каф., к.ф.-м.н., доцент, Шмелева А.Г. |  |  |
| --- | --- | --- | --- |
|  | *(должность, звание, ученая степень, ФИО)* |  | *(подпись руководителя)* |

| **Рецензент** (при наличии) | – |  | – |
| --- | --- | --- | --- |
|  | *(должность, звание, учёная степень)* |  | *(подпись рецензента)* |

Работа представлена к защите « » 2023 г.

Допущен к защите « » 2023 г.

|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА – Российский технологический университет»** |
| **РТУ МИРЭА** |
| Институт кибербезопасности и цифровых технологий |
| *(наименование института, филиала)* |
| Кафедра информатики |
| *(наименование кафедры)* |

Утверждаю

заведующий кафедрой информатики

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Шмелева А.Г.

*Подпись ФИО*

«\_\_\_\_» \_\_\_\_\_\_\_ 20\_\_\_ г.

**ЗАДАНИЕ**

**на выполнение курсовой работы по дисциплине**

**«Основы машинного обучения»**

Студент Нахамкин Константин Дмитриевич Группа БФБО-05-22

**Тема** «Реализация алгоритмов решающих деревьев»

**Исходные данные: Credit Risk Classification Dataset**

**Перечень вопросов, подлежащих разработке, и обязательного графического материала: изучение библиотек NumPy, Pandas, Matplotlib, Sklearn языка программирования Python, методы построения решающих деревьев, решение прикладной задачи анализа датасета, представление результата в виде значений и графиков.**

**Срок предоставления к защите курсового проекта (работы):**  до «\_\_\_» \_\_\_\_\_\_\_\_\_\_ 20\_\_ г.

**Задание на курсовой проект (работу) выдал \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** (\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_)

*Подпись руководителя Ф.И.О. руководителя*

«**01**» **марта**  **2023** г. ***Задание на курсовой проект (работу) получил (\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_)***

*Подпись обучающегося Ф.И.О. исполнителя*.

**Оглавление**

[**Введение 4**](#_heading=h.ho92oy4j8j1s)

[**1. Теоретические основы алгоритмов реализации решающих деревьев 5**](#_heading=h.30j0zll)

[1.1. Основные определения и формулы метода решающих деревьев 5](#_heading=h.1fob9te)

[1.2. Подробное решение модельной задачи 6](#_heading=h.3znysh7)

[**2. Программная реализация решающих деревьев 7**](#_heading=h.2et92p0)

[2.1. Описание датасета и постановка задачи 7](#_heading=h.tyjcwt)

[2.2. Программное решение 8](#_heading=h.3dy6vkm)

[2.3 Анализ полученных данных. Определение погрешности 9](#_heading=h.kwrr83s3w0xa)

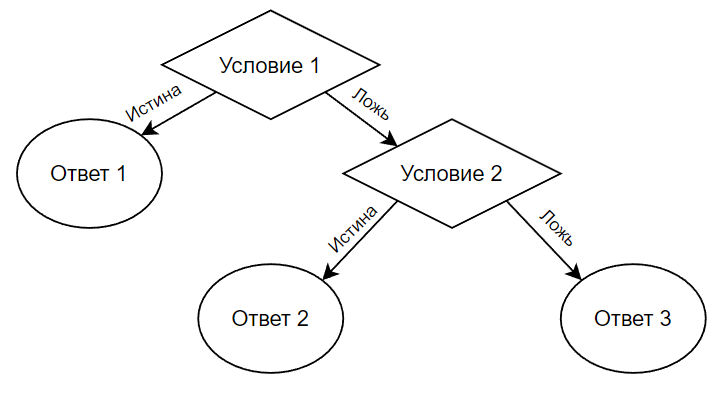
[**Заключение 10**](#_heading=h.1t3h5sf)

[**Список использованных источников 11**](#_heading=h.4d34og8)

[**Приложение 1? 12**](#_heading=h.2s8eyo1)

# Введение

Дисциплина «Основы машинного обучения» направлена на получение навыков создания прикладных программ и реализации алгоритмов на высокоуровневом объектно-ориентированном языке программирования Python. Язык программирования Python является одним из наиболее популярных и востребованных языков, применяемых для решения прикладных задач анализа данных как в крупных компаниях, так и стартапах. Популярность данного языка обеспечивает наличие достаточно большого количества модулей и библиотек, а также бесплатной среды разработки. Деревья решений – популярный и универсальный инструмент интеллектуального анализа данных и предсказательной аналитики, использующийся для решения широкого класса задач. Они представляют из себя графическое представление процесса принятия решений, которое моделирует результаты и предсказывает будущие события. Деревья решений состоят из узлов, представляющих принятие решения, листьев, представляющих окончательные результаты, и веток (ребер графа), соединяющих их друг с другом. Схематическое изображение представлено на рисунке 1.

Рисунок 1 – Пример дерева решений

Деревья решений широко используются в различных областях, включая науку о данных, машинное обучение и искусственный интеллект. Они часто используются при анализе решений для определения наилучшего курса действий с учетом набора возможных результатов и связанных с ними вероятностей. Деревья решений также используются в задачах классификации и регрессии для классификации данных или прогнозирования конкретного результата на основе набора входных переменных.

Построение дерева решений включает в себя процесс, называемый рекурсивным разделением, когда данные разбиваются на более мелкие подмножества на основе наиболее значимого признака. Этот признак выбирается на основе его способности разбивать данные на подмножества, максимально схожие с точки зрения переменной результата. Есть множество алгоритмов выбора признака для следующего ветвления(ID 3, C4.5, CART), но в основном все они основаны на какой-либо формуле понижения энтропии.

Целью данной работы является подробное ознакомление с решающими деревьями и алгоритмами их построения, а также их программная реализация на большом датасете.

# 1. Теоретические основы алгоритмов реализации решающих деревьев

## 1.1. Основные определения и формулы метода решающих деревьев

Текст, текст, текстdsfvmnd sfkjbdsflkbdfbsdfgsdfgsdfbn cvj jdkfbvjdfbgjdf jdfdjgb jkdgbssssss sssssss ssssssssss ssssss sssssssssssssss sssss ssssssssssssssssss sssssssssssssssssssssss.

## 1.2. Подробное решение модельной задачи

Текст, текст, текст

# 2. Программная реализация решающих деревьев

## 2.1. Описание датасета и постановка задачи

pass

## 2.2. Программное решение

pass

## 2.3 Анализ полученных данных. Определение погрешности

pass

# Заключение

pass

# Список использованных источников

1. <https://online-edu.mirea.ru/pluginfile.php?file=%2F1044700%2Fmod_resource%2Fcontent%2F0%2F%D0%A0%D0%B0%D0%B1%D0%BE%D1%87%D0%B0%D1%8F%20%D1%82%D0%B5%D1%82%D1%80%D0%B0%D0%B4%D1%8C%202_%D0%9E%D0%9C%D0%9E_2_%D1%81%D0%B5%D0%BC%D0%B5%D1%81%D1%82%D1%80.pdf>
2. <https://online-edu.mirea.ru/pluginfile.php?file=%2F1047978%2Fmod_resource%2Fcontent%2F0%2F%D0%9B%D0%B5%D0%BA%D1%86%D0%B8%D1%8F%202%20%D0%9F%D0%BE%D1%81%D1%82%D1%80%D0%BE%D0%B5%D0%BD%D0%B8%D0%B5%20%D0%B4%D0%B5%D1%80%D0%B5%D0%B2%D1%8C%D0%B5%D0%B2.pdf>
3. <https://ru.wikipedia.org/wiki/%D0%94%D0%B5%D1%80%D0%B5%D0%B2%D0%BE_%D1%80%D0%B5%D1%88%D0%B5%D0%BD%D0%B8%D0%B9>

# Приложение 1?

Листинг программы?