ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ

(национальный исследовательский университет)» (МАИ)

Направление подготовки: 27.03.05 «Инноватика»

**Лабораторная работа №2**

по дисциплине «Спецглавы высшей математики»

Выполнил:

Студенты гр. М3О-236Б-22

Попов М.А.

Кузьмин Д.Е

Преподаватель:

Александрова С.С.

Москва 2023

Оглавление

[Цель работы. 3](#_Toc153576005)

[Описание теоретической части. 4](#_Toc153576006)

[Реализация метода в коде Python. 5](#_Toc153576007)

[Вывод. 6](#_Toc153576008)

# Цель работы.

Изучить деревья решений и понять алгоритм из работы. Реализовать код из пункта 1 задания, описав его работу. Также описать работу кода из пункта 3 задания. Понять, где применяются деревья решений.

# Описание теоретической части.

Деревья решений — это метод машинного обучения, используемый для прогнозирования результатов на основе входных данных. Основное назначение деревьев решений заключается в классификации или же разделении данных на подгруппы для принятия последовательных решений. Они работают, путем деления набора данных на более мелкие части на основе характеристик, чтобы дойти до конечного вывода. Принципы работы включают рекурсивное разбиение данных и выбор оптимальных признаков для максимизации информативности каждого узла дерева.

Каждый узел в дереве решений представляет собой тест на определенную характеристику данных. Решение о разделении данных принимается с учетом критериев, таких как прирост информации. Процесс построения дерева продолжается, пока не достигнут критерий остановки – энтропия. Энтропия – это критерий, отвечающий за случайность элементов подгруппе. Если он имеет значение 1 – элементы во многом не совпадают, если же 0 – то это означает, что все элементы одинаковые. **Энтропия вычисляется так:** -(p(0) \* log(P(0)) + p(1) \* log(P(1)))

После построения дерево используется для классификации новых данных. Каждый образец проходит через дерево, и на основе тестов в узлах принимается решение о принадлежности к определенному классу.  
Деревья решений могут быть подвержены переобучению, поэтому важно использовать методы обрезки или установку параметров для предотвращения избыточной сложности. Также, они могут быть эффективными инструментами в задачах классификации и регрессии, а также интерпретации данных.

Одним из преимуществ деревьев решений является их интерпретируемость. Легко понять принимаемые решения, так как структура дерева отображает логику процесса принятия решений. Кроме того, деревья способны обрабатывать как категориальные, так и числовые данные без необходимости их предварительной обработки.

Однако деревья решений могут быть чувствительны к изменениям в данных, что может привести к нестабильности модели. Для решения этой проблемы часто используют такие методы, как случайный лес, который комбинирует несколько деревьев для улучшения обобщающей способности модели.

Общий подход к построению и настройке деревьев решений требует внимания к балансу между сложностью модели, ее обобщающей способностью и предотвращением переобучения.

# Реализация метода в коде Python.

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

**import** **matplotlib.pyplot** **as** **plt**

**import** **seaborn** **as** **sns**

df=pd.read\_csv("Diabetes1.csv")

df.head()

x=df.iloc[:,:**2**]

y=df.iloc[:,**2**:]

x.head(**3**)

y.head(**3**)

**from** **sklearn** **import** tree

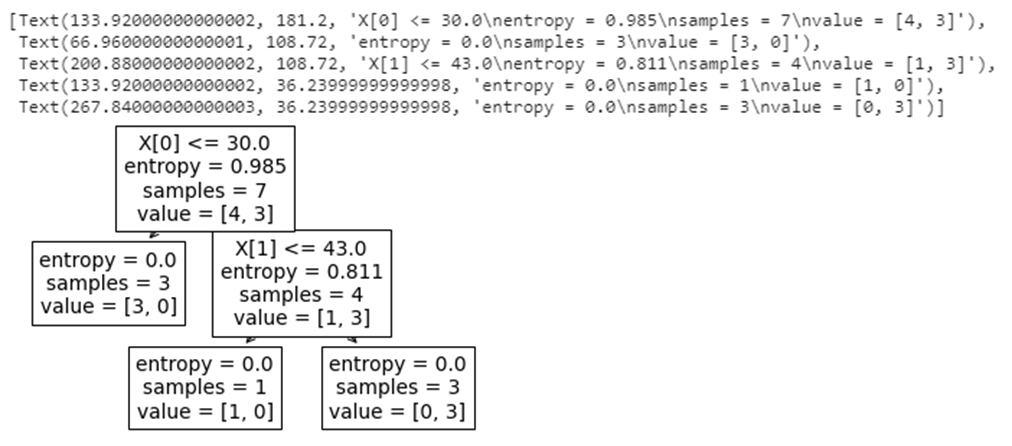
model=tree.DecisionTreeClassifier(criterion="entropy")

model.fit(x,y)

model.score(x,y)

model.predict([[**29**,**47**]])

tree.plot\_tree(model)

Результатом работы программы будет  
  


Данный код выполняет поиск наличия у человека диабета в зависимости от индекса массы тела. Основываясь на показатели интропии, программа выявляет границу bmi, после которой у людей из списка возникает диабет.

Код развивает дерево, пока в каждой из его категорий разделение не будет таково, что энтропия станет равна нулю, то есть все элементы подгруппы будут совпадать по одному значению и удовлетворять условию.

# Вывод.

В ходе выполнения работы был проанализирован и реализован метод дерева принятия решений. Используя язык программирования Python был написан код, описывающий этот метод для некоторых данных из датасета формата .csv. Деревья решений являются одним из самых простых и понятным алгоритмов решения задачи классификации данных и реализации методов машинного обучения. Это достигается простотой работы их узловых точек и небольшим количеством параметров. Из-за этого же они уязвимы к процессы запоминания или же переобучения.