**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ «ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)**

**Кафедра ИС**

**ОТЧЕТ**

**по лабораторной работе №3**

**по дисциплине «Машинное обучение»**

**Тема: Исследование алгоритмов классификации**

Студентка гр. 2373 Никитина М.Ю.

Преподаватель Татчина Я.А.

Санкт-Петербург

2024

Задание:

1. Необходимо оценить и сравнить результаты классификации, используя следующие

алгоритмы классификации:

\* kNN

\* дерево решений

2. Сравните полученные результаты с помощью различных метрик оценки качества:

Accuracy

Presicion, Recall, F-measure

ROC

3. Объяснить полученные результаты.

Для выполнения данной лабораторной работы была взята база данных Music Education Dataset, скаченная с сайта <https://www.kaggle.com/datasets/ziya07/music-education-performance-data>.

В данном датасете представлена информация о том, как сильно влияет музыка на умственную деятельность студентов. Все данные реальны.

Датасет включает в се6я следующие атрибуты:

Student\_ID - номер студента (String)

Age - возраст студента (Integer)

Gender - пол сутдента (String)

Class\_Level - уровень музыкального образования (String)

Accuracy - точность воспроизведения музыки студентом (Float)

Rhythm - способность студента удерживать ритм (Float)

Tempo - темп выступления студента (Float)

Pitch\_Accuracy - оценка точности подачи в выступлении студента (Float) Duration - время игры студента (Float)

Volume - громкость выступления студента (Float)

Heart\_Rate - пульс студента во время урока (Float)

Blood\_Pressure - давление студента во время урока (Float)

Stress\_Level - уровень стресса студента (от 1 до 10) (Integer) Engagement\_Level - уровень вовлеченности студента (Integer)

Focus\_Time - время фокусировки студента на задании (Float) Behavioral\_Patterns - оценка отвлеченности студента (Integer) Performance\_Score - оценка выступления студента (Float)

Skill\_Development - оценка развития музыкальных навыков студента (Integer)

Engagement\_Score - оценка врвлеченности студента (Float)

Lesson\_Type - тип урока (String)

Instrument\_Type - вид инструмента (String)

В качестве целевого класса был выбран столбец Gender (пол студента), где 1 – женщина, 0 – мужчина.

1. Оценим и сравним результаты классификации с помощью дерева решений и knn:

Мы разделили выборку на обучающую и отложенную. В результате, для дерева решений на обучающей выборке средняя доля верных ответов составила ровно 0.5, а после того, как мы обучили модель, доля стала 0.61428571.

Для knn на обучающей выборке средняя доля верных ответов составила 0.5142857142857142, а после обучения модели доля стала 0.58571428571.

Сравнив полученные результаты, можно сделать вывод, что после обучения моделей дерево решений лучше справилось со своей задачей.

1. Сравним полученные результаты:
   1. Accuracy

Для дерева решений: 0.43333333333333335

Для knn: 0.5

* 1. Precision

Для дерева решений: 0.46153846153846156

Для knn: 0.5

* 1. Recall

Для дерева решений: 0.8

Для knn: 0.6666666666666666

* 1. F-measure

Для дерева решений: 0.5853658536585366

Для knn: 0.5714285714285714

* 1. ROC

Изображение выглядит как линия, текст, диаграмма, График

Автоматически созданное описаниеДля дерева решений:

Изображение выглядит как линия, текст, диаграмма, График

Автоматически созданное описаниеДля knn:

1. Проанализировав полученные результаты, можно сказать, что knn лучше справилось с заданием в таких метриках, как Accuracy, Precision и ROC. В остальных же случаях, дерево решений дало более верный ответ.

Поэтому здесь нельзя однозначно сказать, какой из методов работает лучше, так как результаты показали, что в одном случае лучше работает knn, а в других дерево решений.

Вывод:

В данной работе мы изучили два метода классификации: knn и дерево решений. Как уже было сказано выше, каждый метод хорош в определенной метрике, так что для дальнейшей классификации данных из этого датасета лучше применять оба метода.