**МИНОБРНАУКИ РОССИИ**

**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 | Маркова З.А. |
| Преподаватель | Татчина Я.А. |

Санкт-Петербург

2024

**Цель работы**: оценить и сравнить результаты классификации, используя алгоритмы kNN и дерево решений, сравнить полученные результаты с помощью метрик качества и объяснить их.

1. **Краткое описание датасета**

Мною был выбран новый датасет Weather Forecast Dataset, так как прошлый не подходил для задачи классификации.

Этот датасет содержит данные с наблюдениями за погодой.

Атрибуты:

1. Temperature – Температура. Тип данных числовой.
2. Humidity – Влажность. Тип данных числовой.
3. Wind\_Speed – Скорость ветра. Тип данных числовой.
4. Cloud\_Cover – Покрытие облаками. Тип данных числовой.
5. Pressure – Давление. Тип данных числовой.
6. Rain – Дождь. (Наличие или отсутствие).
7. **Целевой класс**

В качестве целевого класса был выбран атрибут «Rain», его значение были изменены на {rain=1, no rain=0}. Датасет не сбалансирован: количество элементов равных 1 - 314, равных 0 – 2186.

1. **Оценить и сравнить результаты классификации**

**3.1. Дерево решений**

Средняя оценка качества составила: 0.9948571428571429.

Затем для дерева решений были настроены лучшие значения параметров: 'max\_depth': 6, 'max\_features': 4.

После обучения лучшее качество составило: 0.9965714285714287

**3.2. kNN**

Средняя оценка качества составила: 0.9697142857142858.

Для Метода ближайших соседей было настроено лучшее значение параметра 'knn\_\_n\_neighbors': 9.

После обучения лучшее качество составило: 0.9662857142857142

Дерево решений справляется немного лучше, чем kNN. Показывая верный результат в 99%, против 96% у kNN.

4. Сравнение полученных результаты с помощью различных метрик оценки качества:

4.1. Метрики оценки качества Дерева решений

Accuracy: 0.9986666666666667

Precision: 1.0

Recall: 0.9896907216494846

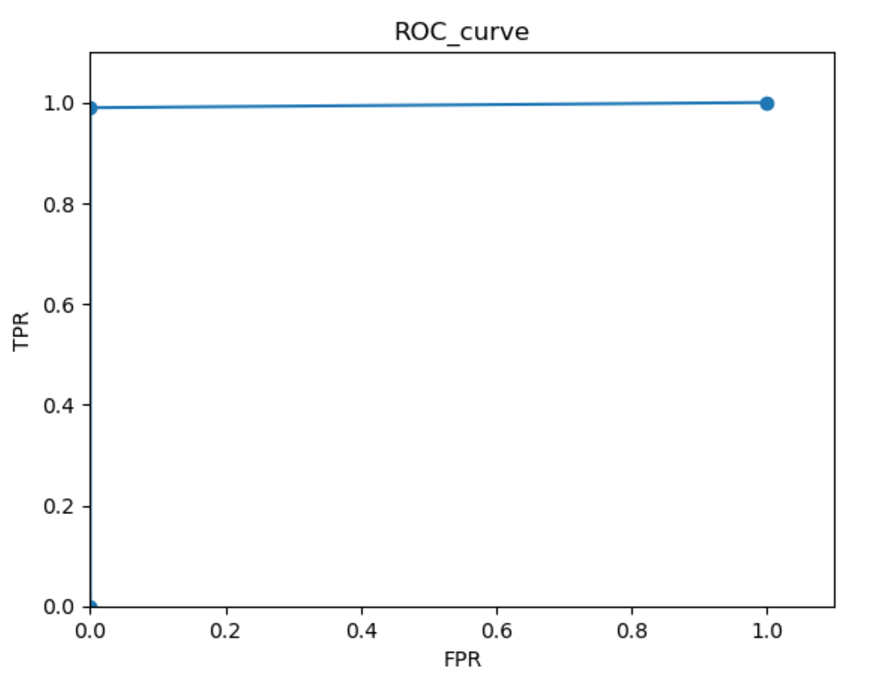
F1: 0.9948186528497409

Рис. 1 Оценка качества ROC

4.2. Метрики оценки качества kNN

Accuracy: 0.9693333333333334

Precision: 0.9111111111111111

Recall: 0.845360824742268

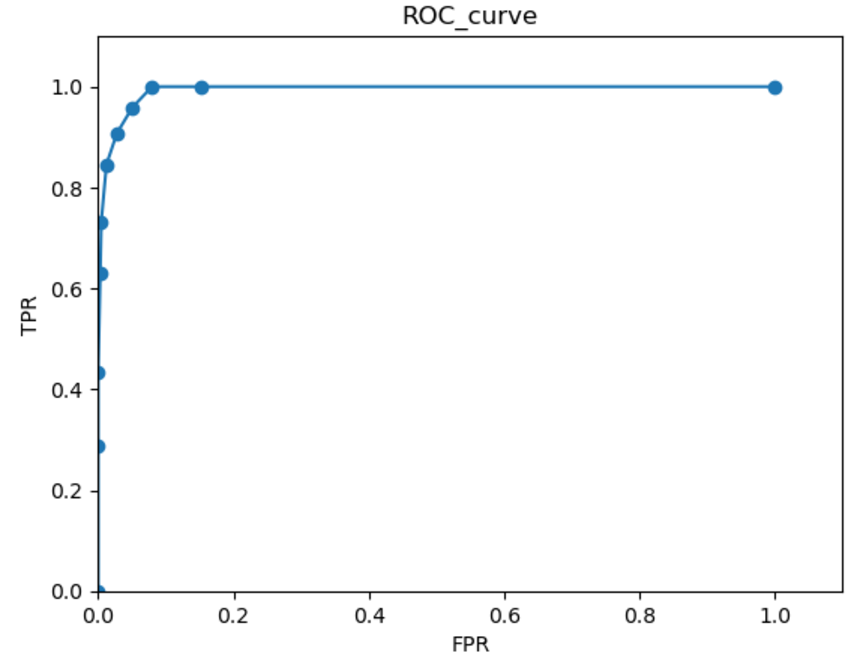
F1: 0.8770053475935828

Рис. 2 Оценка качества kNN

По получившимся значениям метрик оценки качества видно, что Дерево Решений справляется лучше с задачей классификации, чем kNN, показывая результат близкий к точности 100%.

Также по графикам ROC-кривой видно, что Дерево Решений более хороший классификатор, так как получается максимальный TRP при минимальном FRP, а площадь под кривой близка к 1. Значит Дерево Решений правильно определяет класс почти для всех значений.

**Вывод**

В данной лабораторной работе были рассмотрены алгоритмы классификации такие как: Дерево Решений и Метод ближайших соседей (kNN) для классификации данных датасета Weather Forecast Dataset.

В качестве целевого класса был выбран атрибут “Rain”, который показывает наличие или отсутствие дождя.

Благодаря метрикам оценки качества было выявлено, что Дерево Решений лучше справляется с задачей классификации, чем kNN, правильно определяя класс почти во всех случаях.