**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра информационных систем**

отчет

**по лабораторной работе №3**

**по дисциплине «Машинное обучение»**

Тема: Исследование алгоритмов классификации

|  |  |  |
| --- | --- | --- |
| Студент гр. 2372 |  | Михеева Э. Д. |
| Преподаватель |  | Татчина Я. А. |

Санкт-Петербург

2024

**Цель работы**

Изучение задачи и методов классификации, получение навыков применения методов kNN (k ближайших соседей) и дерева решений.

**Постановка задачи**

1. Оценить, какой атрибут выступит в роли целевого класса (метки);
2. Оценить сбалансированность классов;
3. Применить методы классификации алгоритмами kNN, Дерево решений;
4. Сравнить полученные результаты с помощью различных метрик оценки качества: accuracy, precision, recall, f1, ROC;
5. Проанализировать полученные результаты.

**Выполнение работы**

Продолжая работу над набором данных о ценовых категориях мобильных устройств, выделен целевой класс – “price\_range” (ценовая категория) и разделен набор на обучающую и тестовые выборки с помощью библиотеки sklearn.

Методом локтя подобрано наилучшее число соседей для алгоритма kNN (k = 15), график показан на рисунке 1.

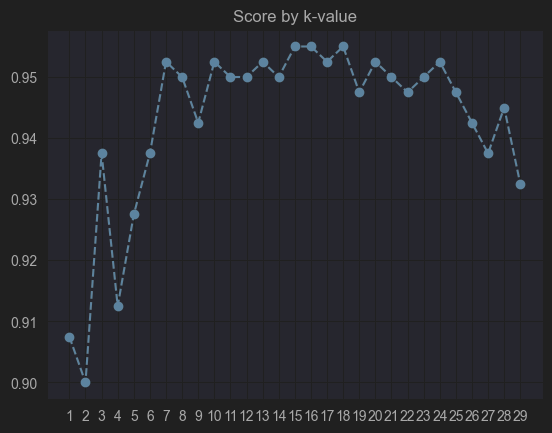


Рисунок . Подбор значения k для kNN

Далее была обучена модель и описана функция отображения необходимых метрик. Для оценки результатов применяется матрица корреляции, оценки доли правильных ответов (accuracy), точности (precision), полноты (recall), f1- и ROC-меры.

Аналогично, была обучена модель дерева решений и проведен анализ полученных результатов на основе метрик (рис. 2).



Рисунок . Метрики kNN и DT

Благодаря сбалансированности набора данных, такие простые методы, как k-ближайших соседей и дерево решений, показали хорошие результаты. Стоит отметить, что в данной ситуации kNN выигрывает почти на 15%. Это объясняется тем, что классическое дерево решений работает лучшим образом именно с бинарными метками, а в исследуемом наборе 4 класса цены.

**Выводы**

В ходе работы была проведена классификация методами k-ближайших соседей и дерева решений. Были обучены соответствующие модели, оценены и сравнены результаты, получены высокие метрические оценки.