МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение

высшего образования

«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра «Измерительно-вычислительные комплексы»

«Методы искусственного интеллекта»

Отчёт по лабораторной работе №5

Выполнил:

студент группы ИСТбд-42

Осипов К. С.

Проверил:

доцент кафедры ИВК, к.т.н.

Шишкин В. В.

Ульяновск

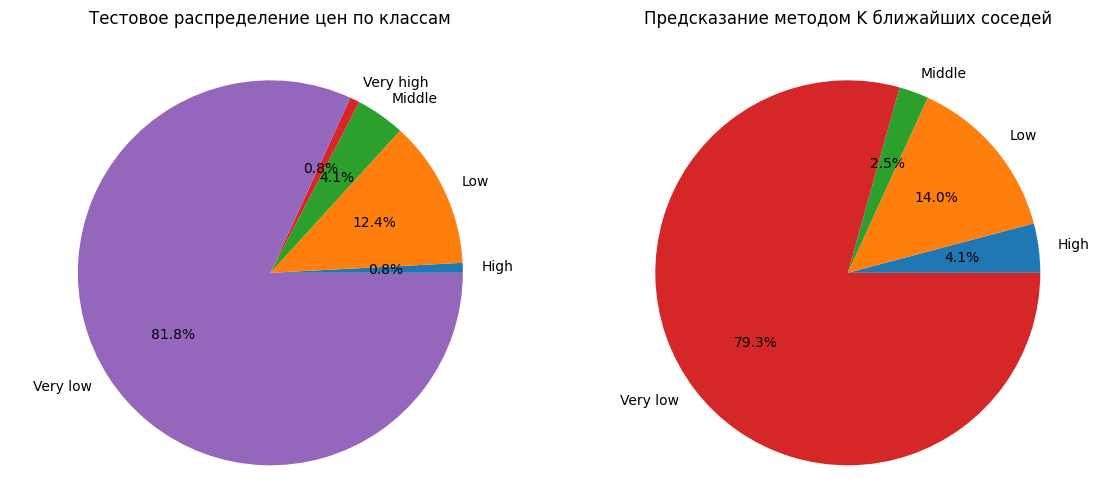
2022

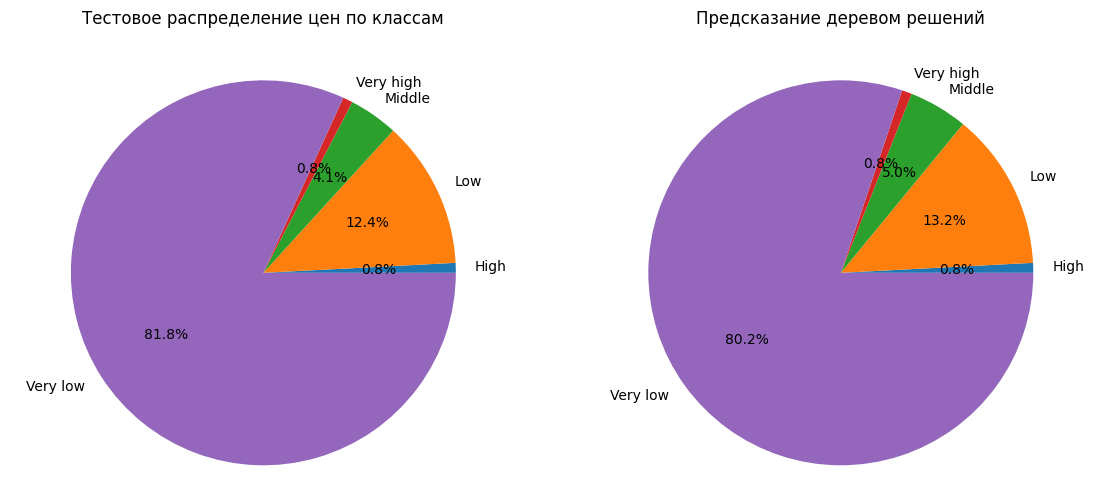
**Определение целевого столбца и набора признаков**

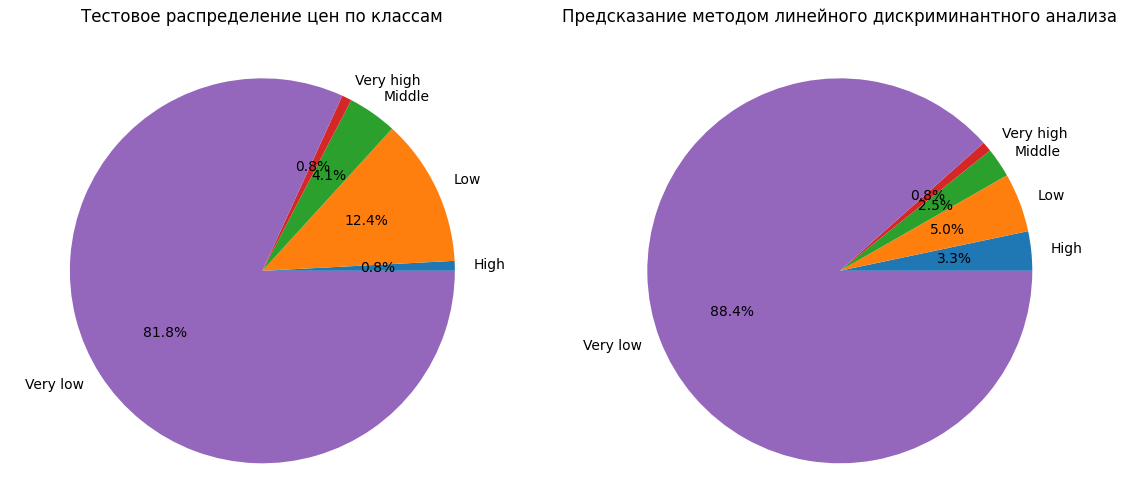
В исходном наборе данных не было целевого столбца, поэтому он был добавлен программно на основе уровня цен на автомобили (листинг ниже). Столбец Present\_Price (текущая цена) был удален, т.к. рассматривается только цена продажи (Selling\_Price). Все остальные столбцы использовались в качестве набора признаков для классификации.

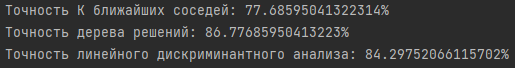
price\_classes = ['Very low', 'Low', 'Middle', 'High', 'Very high']  
  
# Добавление столбца для классификации цен  
for i in range(len(cars\_dataset)):  
 if 0.0 <= cars\_dataset['Selling\_Price'][i] < 7.0:  
 cars\_dataset['Price\_Class'][i] = price\_classes[0]  
 continue  
 if 7.0 <= cars\_dataset['Selling\_Price'][i] < 14.0:  
 cars\_dataset['Price\_Class'][i] = price\_classes[1]  
 continue  
 if 14.0 <= cars\_dataset['Selling\_Price'][i] < 21.0:  
 cars\_dataset['Price\_Class'][i] = price\_classes[2]  
 continue  
 if 21.0 <= cars\_dataset['Selling\_Price'][i] < 28.0:  
 cars\_dataset['Price\_Class'][i] = price\_classes[3]  
 continue  
 if 28.0 <= cars\_dataset['Selling\_Price'][i] <= 35.0:  
 cars\_dataset['Price\_Class'][i] = price\_classes[4]  
 continue

**Сравнение точности предсказаний различных классификаторов**









**Выводы**

Из круговых диаграмм и вывода программы видно, что самым точным оказалась классификация при помощи дерева решений (86,78%), затем идет линейный дискриминантный анализ (84,30%), и наименее точный метод K ближайших соседей (77,69%).

К тому же, метод KNN не выделил отдельный класс цен «Very high». Это связано с тем, что в наборе данных автомобили с высокими ценами (классы «High» и «Very high») встречаются горазда реже остальных. Тестовое распределение показывает, что на 2 этих класса приходится всего около 1,6% от всех автомобилей.