|  |  |
| --- | --- |
|  | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

**Факультет «Информатика и системы управления»**

**Кафедра ИУ5 «Системы обработки информации и управления»**

Дисциплина «Машинное обучение»

ОТЧЕТ

ПО ЛАБОРАТОРНОЙ РАБОТЕ 3

Тема: «Прогноз погоды»

Студент: Акулова А.А.

Группа ИУ5-65Б

Преподаватель: Гапанюк Ю.Е.

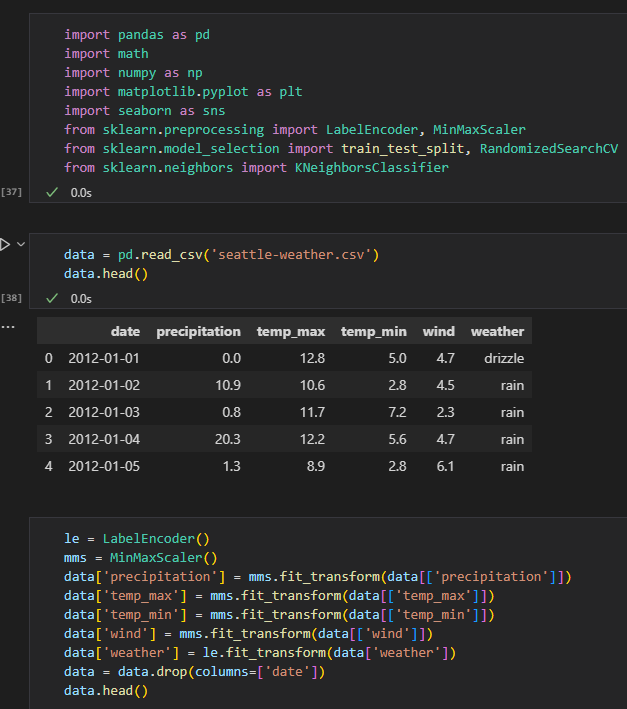
2022г.

**Задание:**

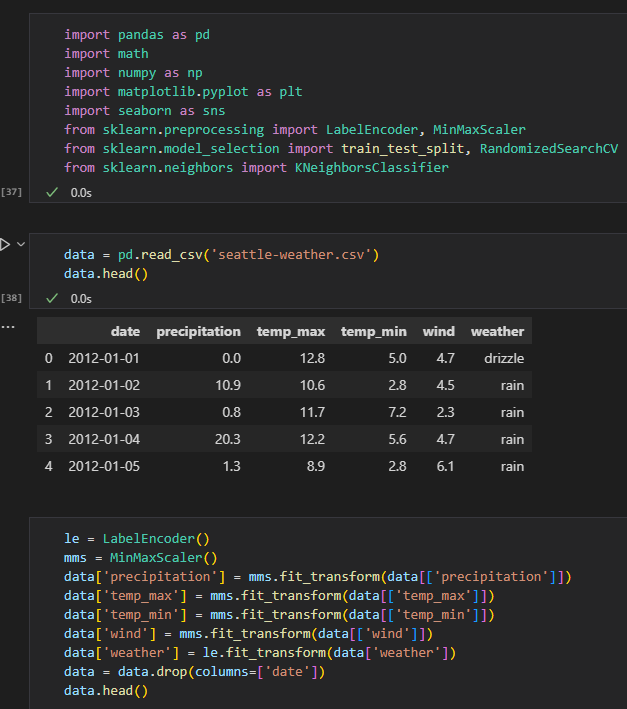
1. Выберите набор данных (датасет) для решения задачи классификации или регрессии.
2. С использованием метода train\_test\_split разделите выборку на обучающую и тестовую.
3. Обучите модель ближайших соседей для произвольно заданного гиперпараметра K. Оцените качество модели с помощью подходящих для задачи метрик.
4. Произведите подбор гиперпараметра K с использованием GridSearchCV и/или RandomizedSearchCV и кросс-валидации, оцените качество оптимальной модели. Желательно использование нескольких стратегий кросс-валидации.
5. Сравните метрики качества исходной и оптимальной моделей.

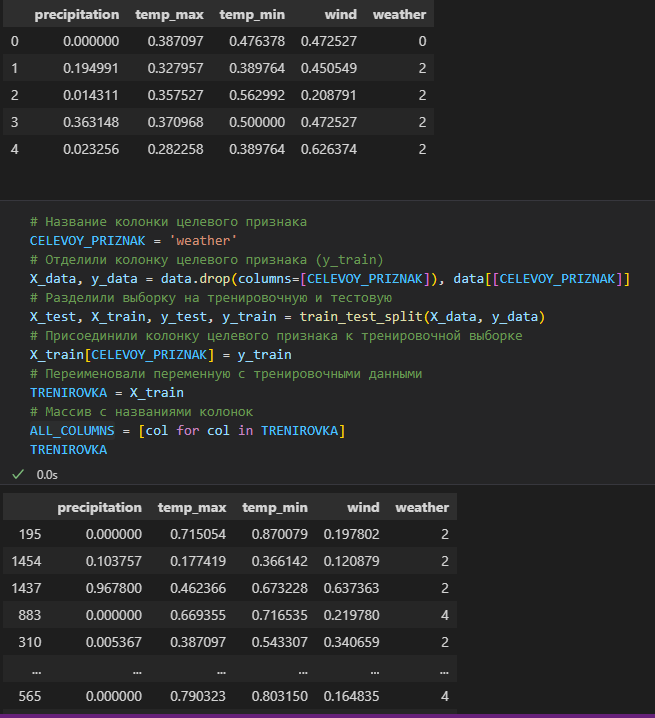
**Текст программы и графики**

Импотируем датасет

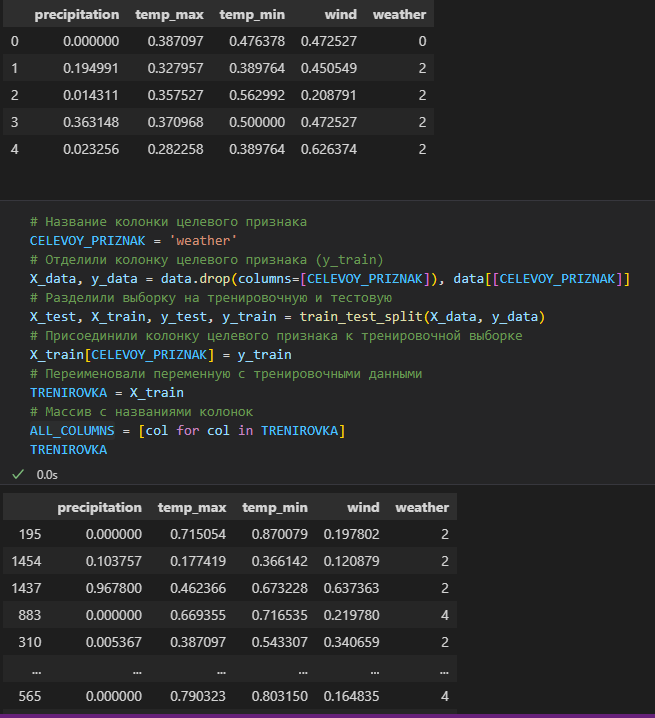


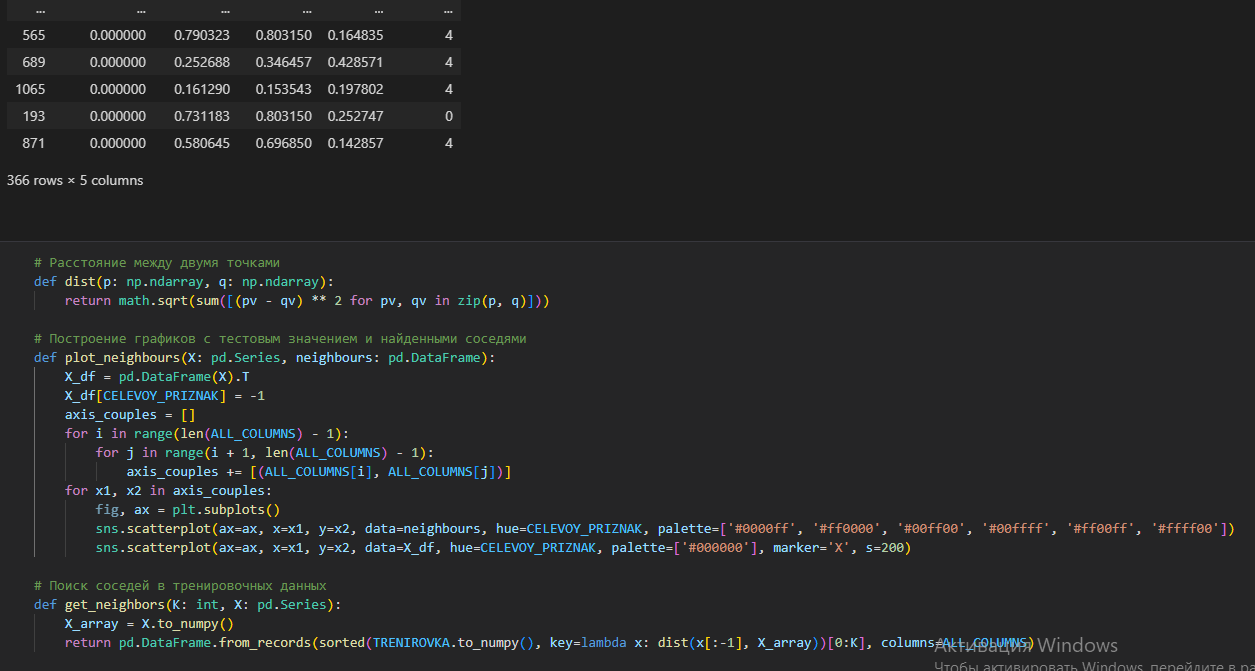
Удаление лишних столбцов, кодирование категориальных признаков, транспонирование столбцов, по которым будет вестись предсказание.



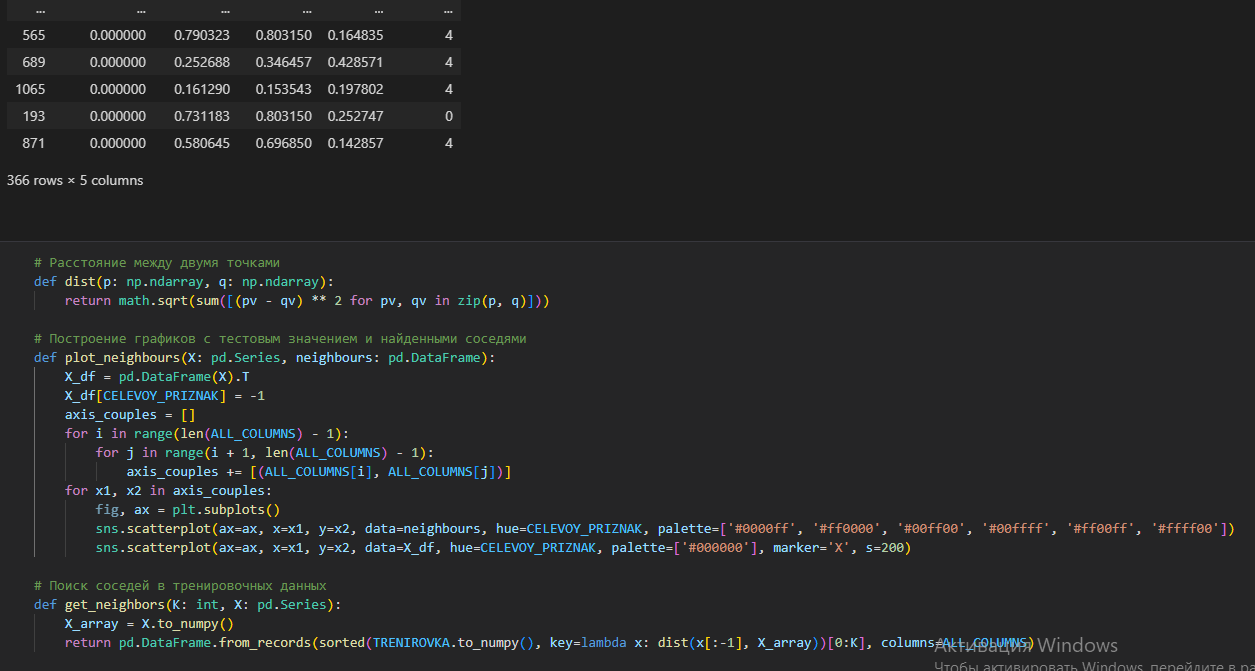


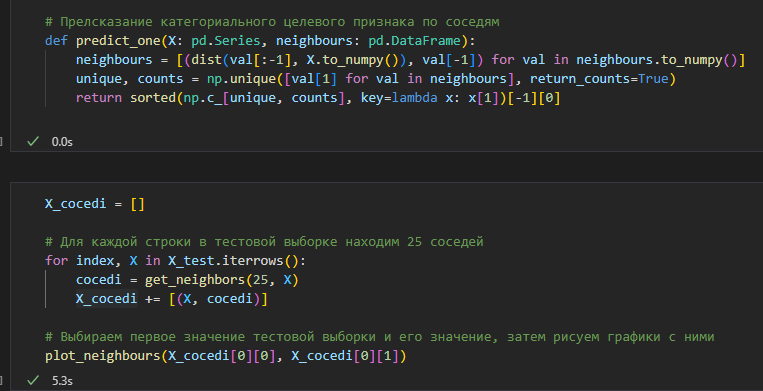
Разделение выборки на тестовую и тренировочную.



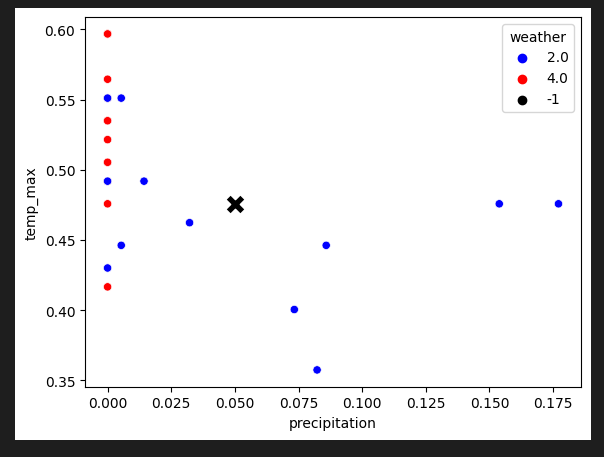


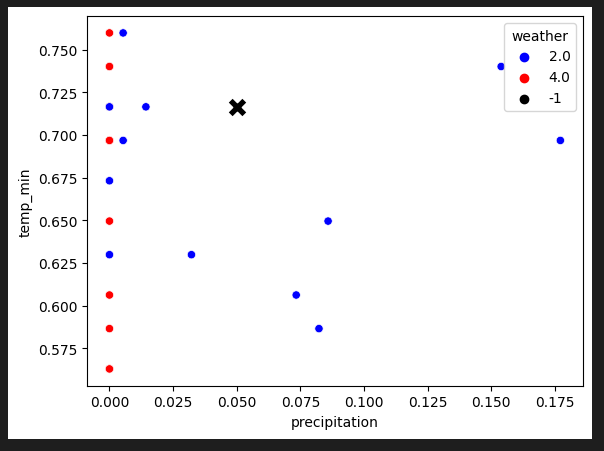
Написание алгоритма по выборке случайной точки для тренировочной выборки, поиск 25 соседей, обучение предсказыванию.

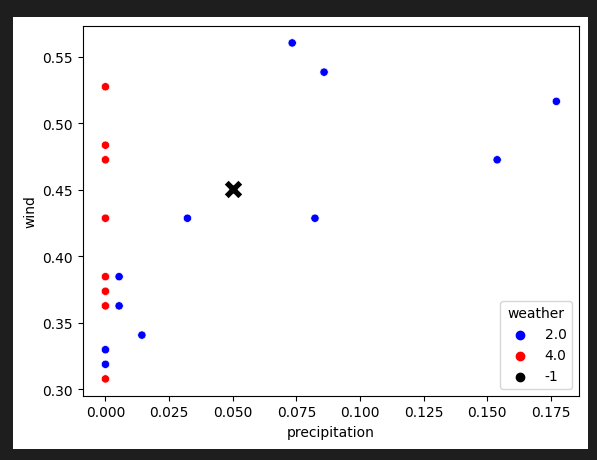


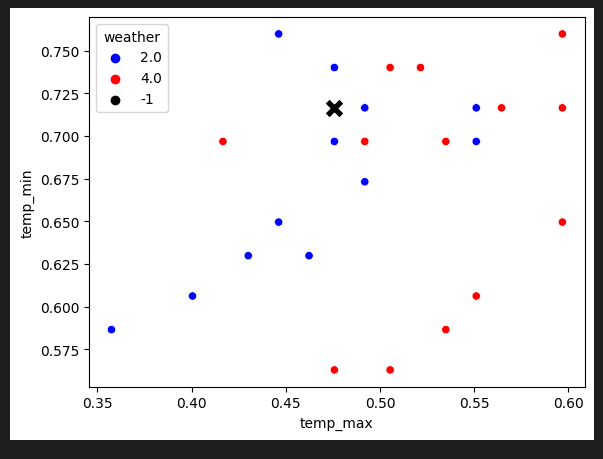


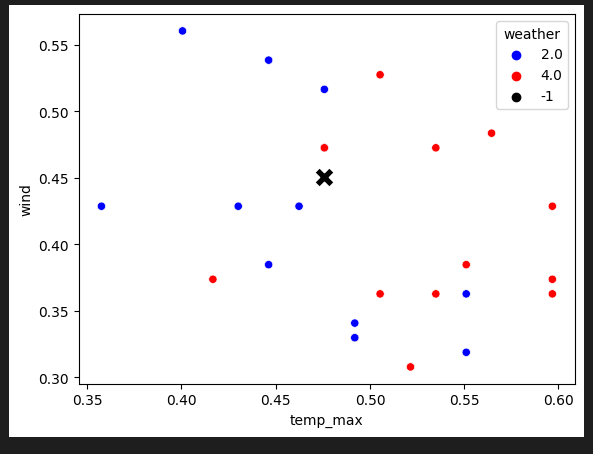
Так как предсказания проводятся по четырем колонкам, то имеем четырехмерное изображение. Построим графики для всех 6 возможных пар осей.

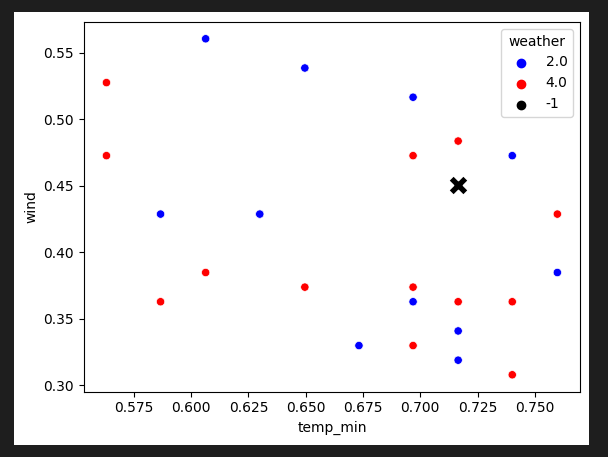




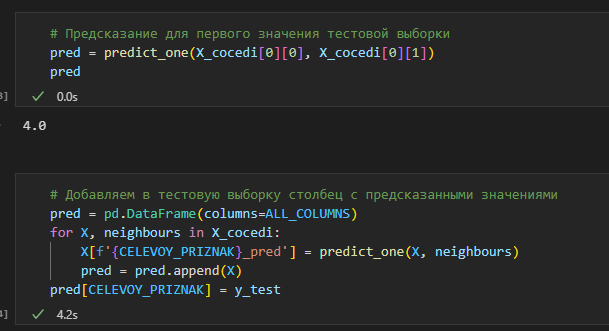


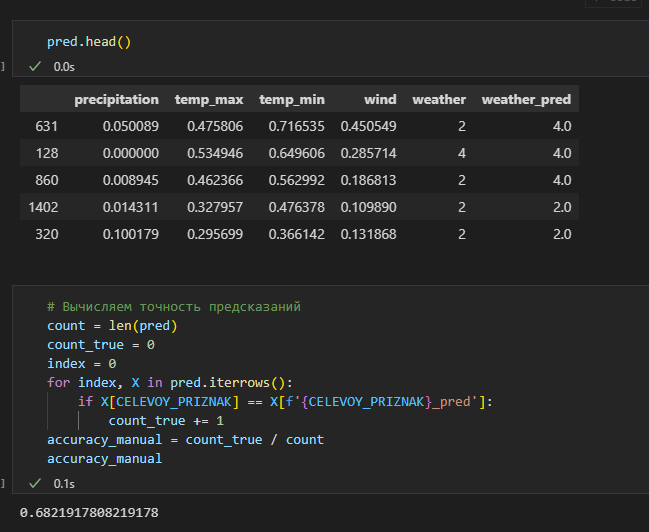




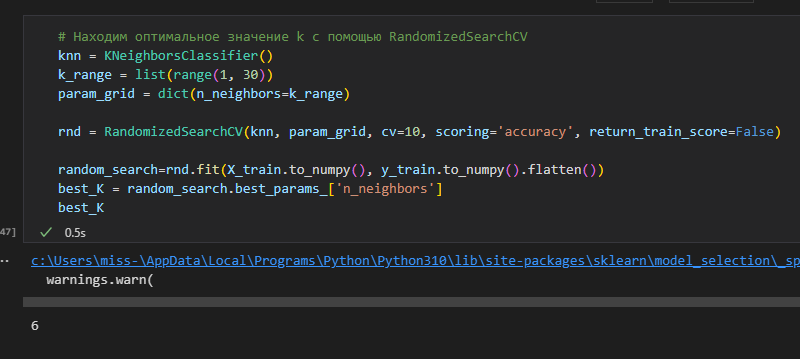


Предсказание для первого значения и измерение точности предсказаний от тренировочной выборки.





Определение количества необходимых соседей для более точных результатов.



Предсказание для первого значения и измерение точности предсказаний от тестовой выборки.

