Московский государственный технический университет им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Автоматизированные системы обработки информации и управления»



**Отчет**

**Лабораторная работа № 4**

**«Подготовка обучающей и тестовой выборки, кросс-валидация и подбор гиперпараметров на примере метода ближайших соседей»**

**По курсу «Технологии машинного обучения»**

**ИСПОЛНИТЕЛЬ:**

Елизаров Олег

Группа ИУ5-63

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

"\_\_"\_\_\_\_\_\_\_\_\_\_\_2020 г.

**ПРЕПОДАВАТЕЛЬ:**

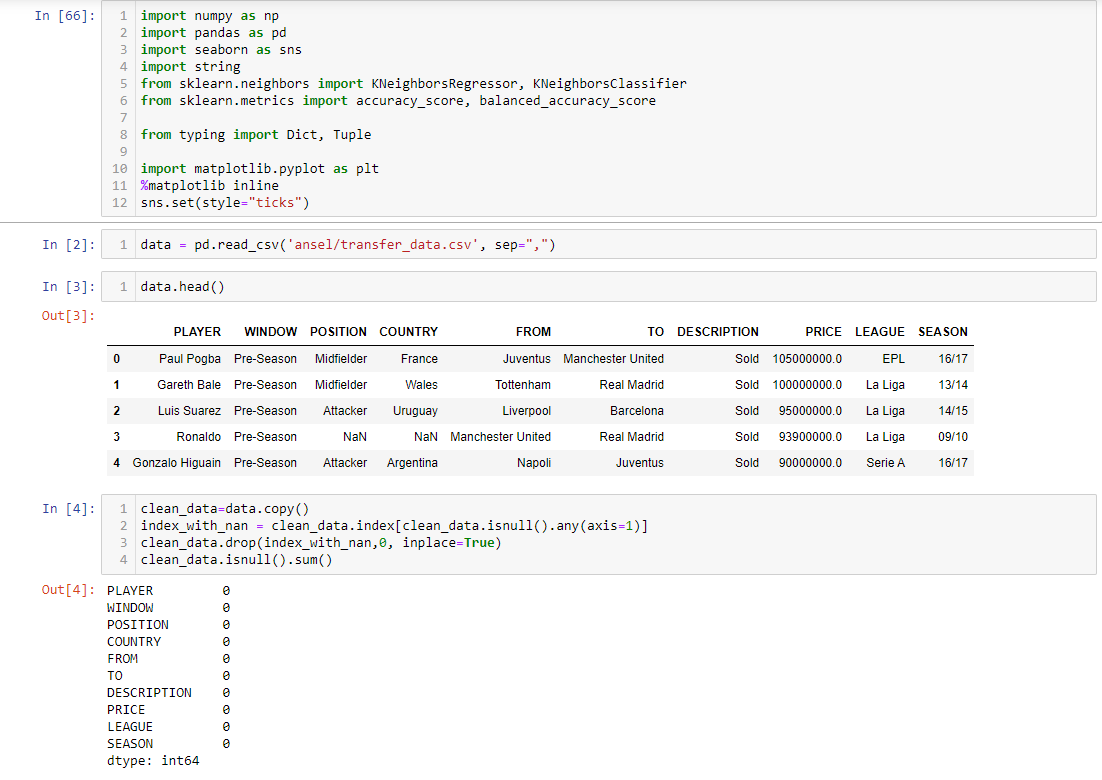
Гапанюк Ю.Е.

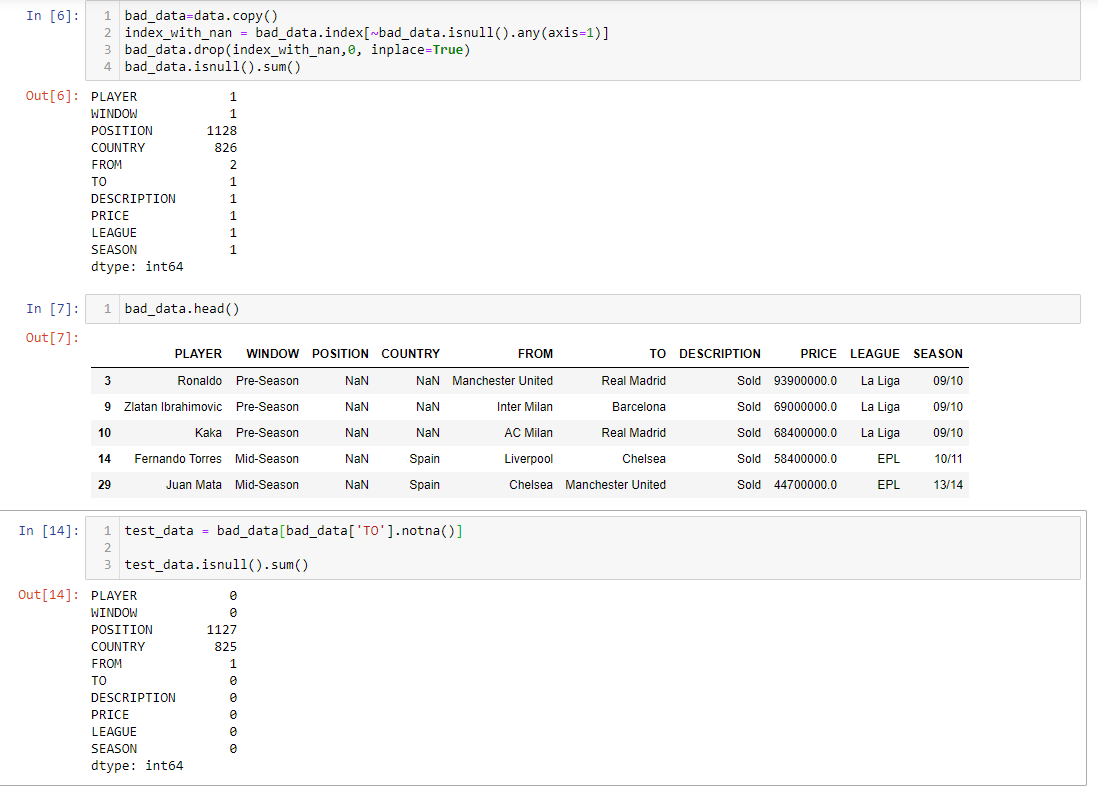
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

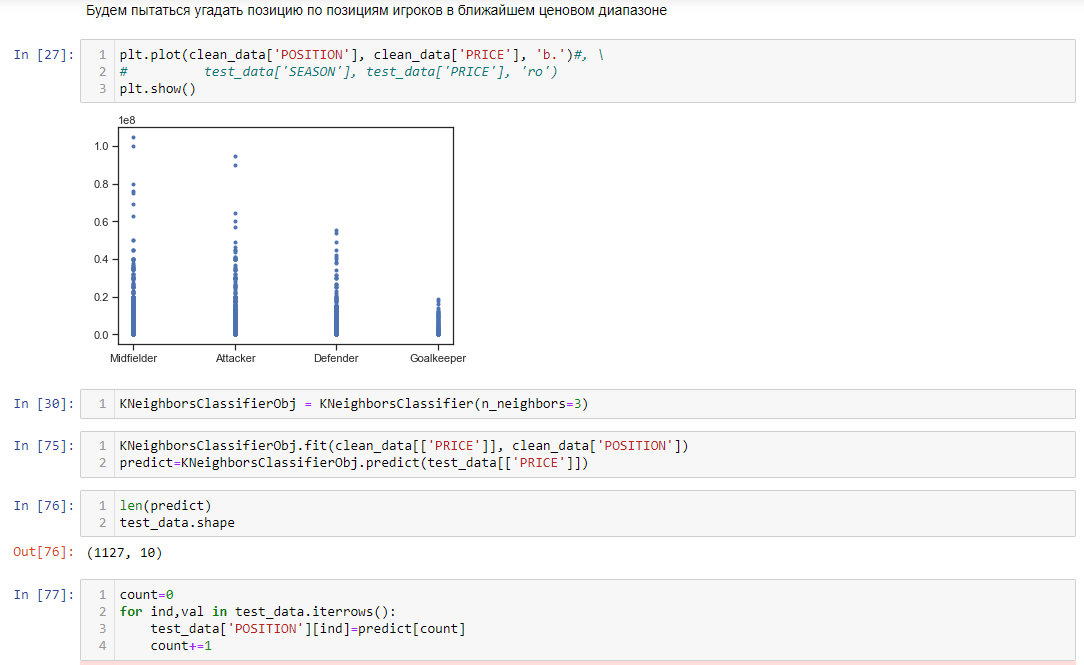
"\_\_"\_\_\_\_\_\_\_\_\_\_\_2020 г.

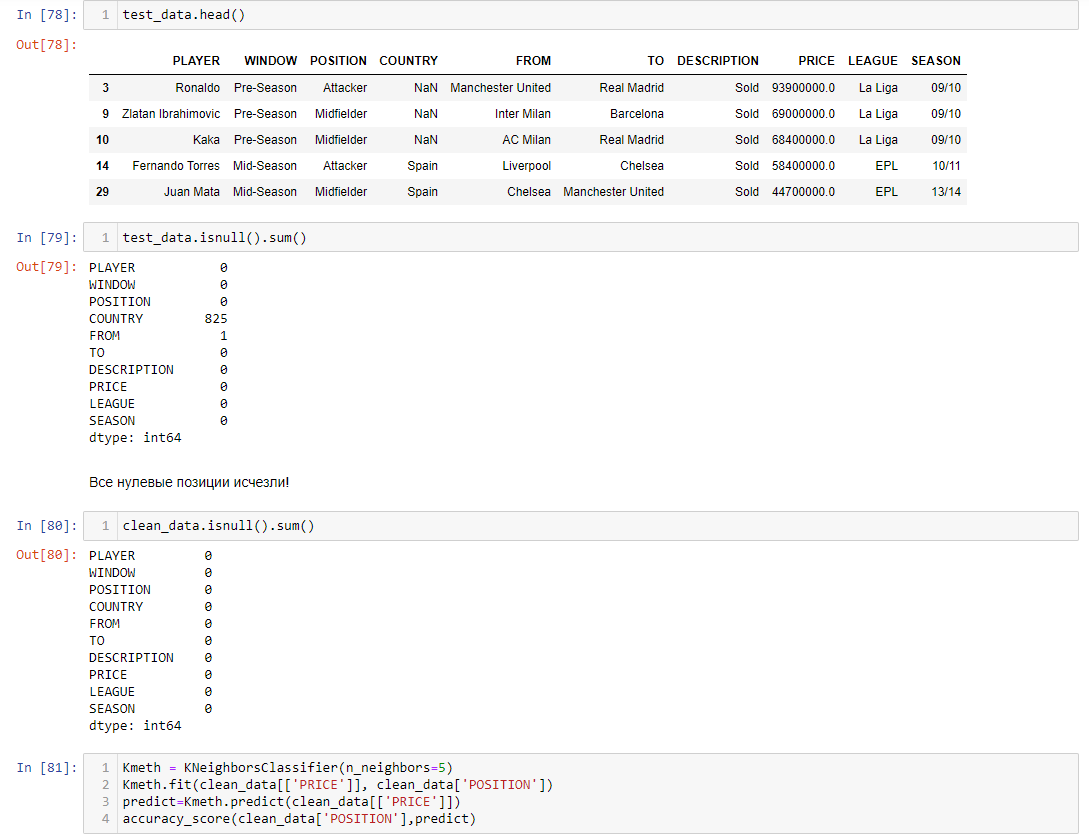
Москва2020

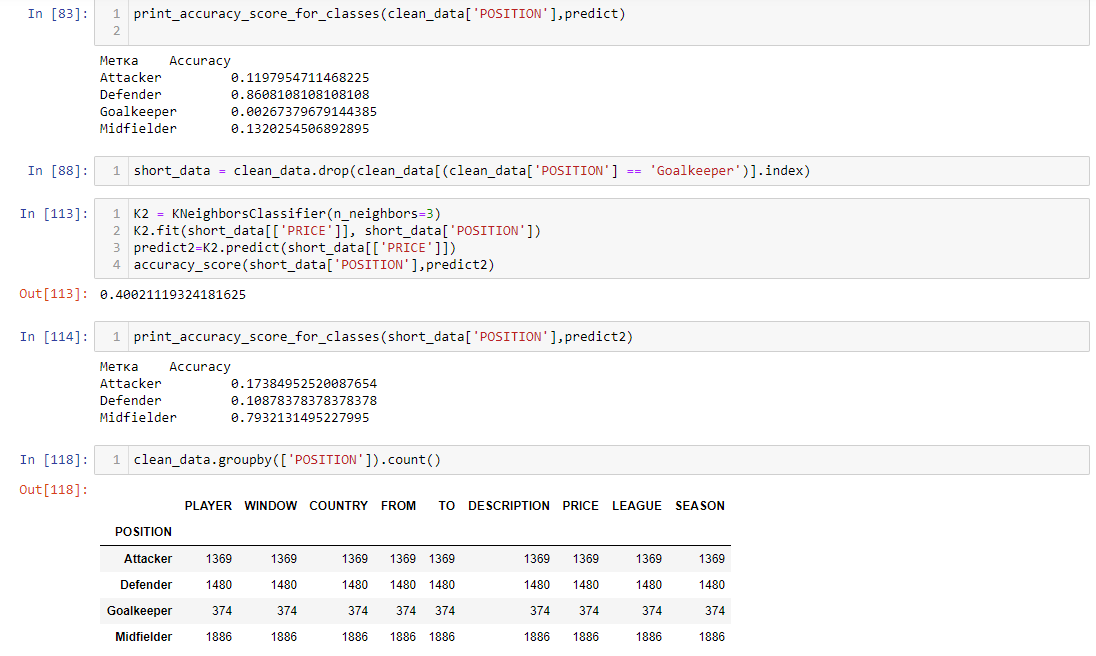
1. **Цели работы:**Изучение способов предварительной обработки данных для дальнейшего формирования моделей.
2. **Задание**:
3. Выберите набор данных (датасет) для решения задачи классификации или регрессии.
4. С использованием метода train\_test\_split разделите выборку на обучающую и тестовую.
5. Обучите модель ближайших соседей для произвольно заданного гиперпараметра K. Оцените качество модели с помощью подходящих для задачи метрик.
6. Постройте модель и оцените качество модели с использованием кросс-валидации.
7. Произведите подбор гиперпараметра K с использованием GridSearchCV и кросс-валидации.
8. **Выполнение работы:**

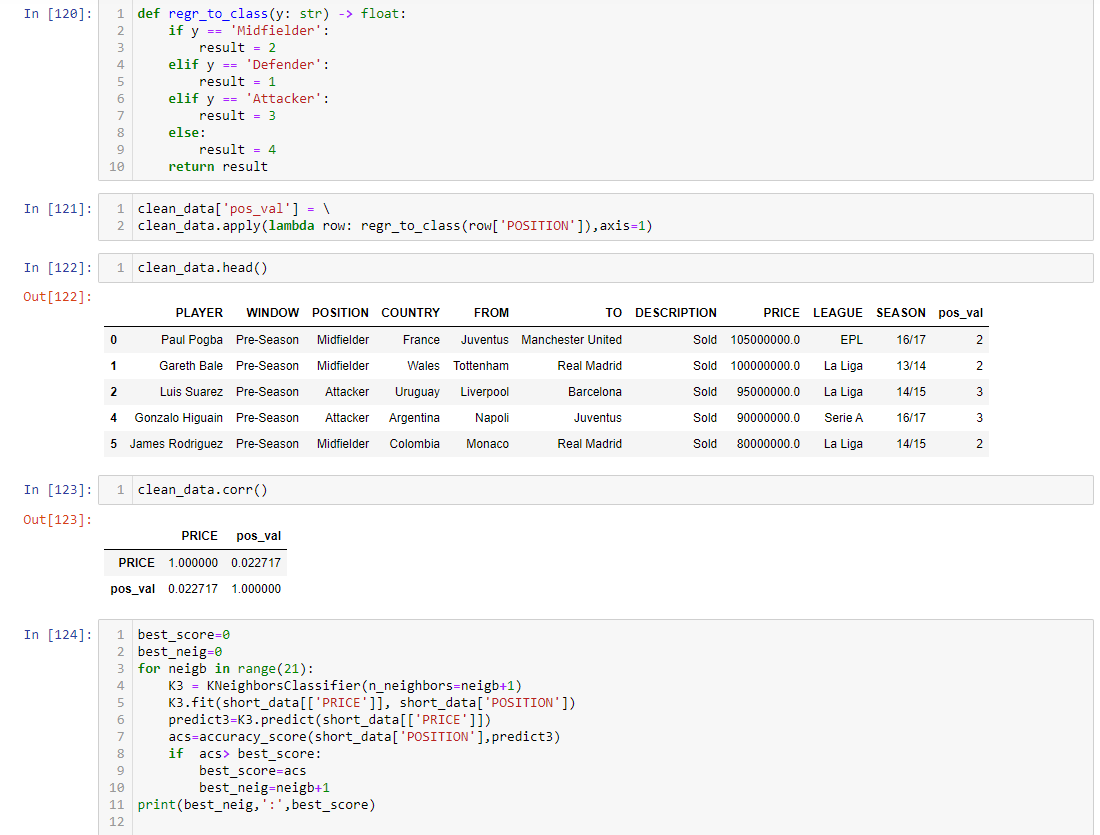












1. **Заключение:**

Итак, мы пытались установить зависимость между позицией игрока и ценой, за которую был совершен его трансфер. Первоначально мы разделили наши данные на две части: позиция присутствует, а во второй отсутствует. Обучившись на первой выборке мы заполнили пустые значения во второй. Но проведя множество анализов, мы выяснили, что хоть и влияние определенно имеется, но оно не столь значимо ,чтоб можно было только по цену судить о позиции. Мы заметили, что вратарей в выборке изначально мало, так что убрали их и произвели подборку лучшего результата и лучшего количества соседей. Но результаты говорят, что мы угадываем примерно 4 из 10 игроков, что, конечно, не полный провал, но и явно не успех. Необходимы дальнейшие исследования и более богатая выборка.