**Московский государственный технический**

**университет им. Н.Э. Баумана**

Факультет «Информатика с системы управления»

Кафедра ИУ5 «Системы обработки информации и управления»

Курс «Технологии машинного обучения»

Отчёт по лабораторной работе №3

# «[Подготовка обучающей и тестовой выборки, кросс-валидация и подбор гиперпараметров на примере метода ближайших соседей](https://github.com/ugapanyuk/courses_current/wiki/LAB_TMO__KNN)[.](https://github.com/ugapanyuk/courses_current/wiki/LAB_TMO__MISSING)»

Выполнил: Проверил:

студент группы РТ5-61Б преподаватель каф. ИУ5

Мицкевич В.Б. Гапанюк Ю.Е.

Подпись и дата: Подпись и дата:

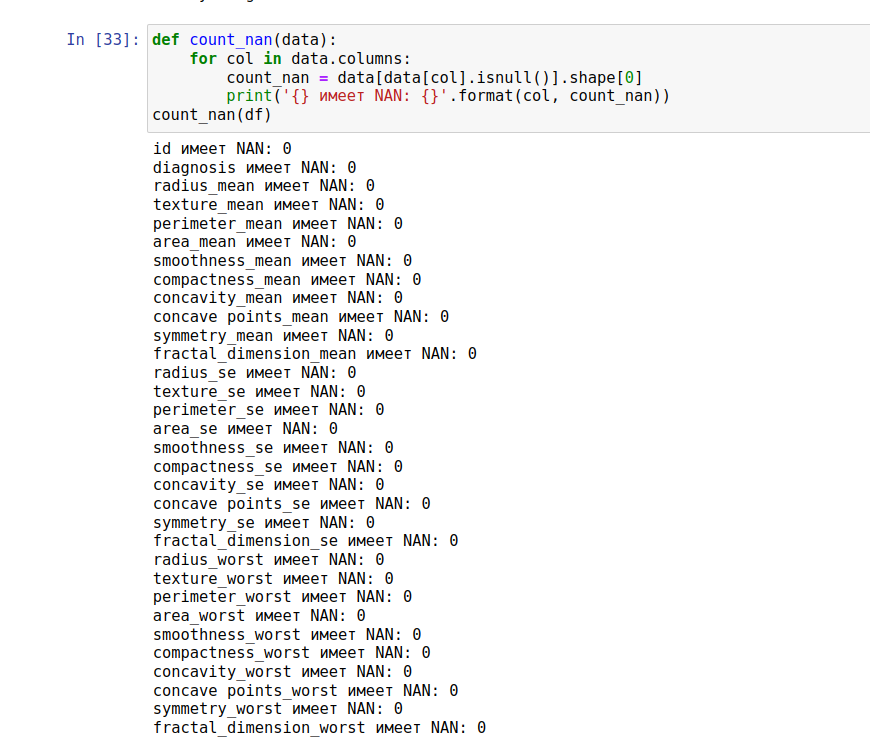
Москва, 2023 г.

## **Задание:**

Выберите набор данных (датасет) для решения задачи классификации или регрессии. В случае необходимости проведите удаление или заполнение пропусков и кодирование категориальных признаков. С использованием метода train\_test\_split разделите выборку на обучающую и тестовую. Обучите модель ближайших соседей для произвольно заданного гиперпараметра K. Оцените качество модели с помощью подходящих для задачи метрик. Произведите подбор гиперпараметра K с использованием GridSearchCV и RandomizedSearchCV и кросс-валидации, оцените качество оптимальной модели. Используйте не менее двух стратегий кросс-валидации. Сравните метрики качества исходной и оптимальной моделей.

В качестве датасета для рабораторной работы был взят датасет из kaggle для модели ближайших соседей

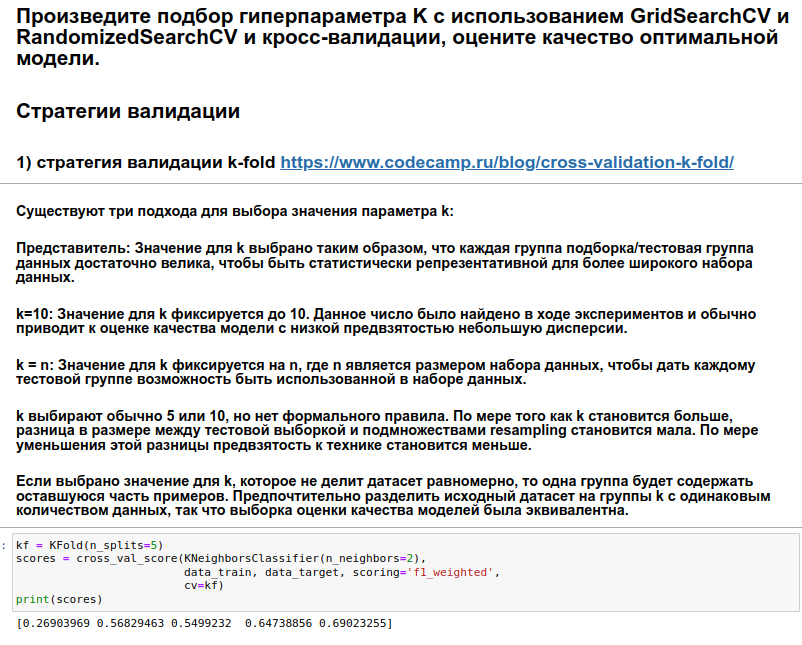
Датасет не имеет пропусков. Из категориальных признаков только колонка „diagnosis“, который будет являться классификацией для модели МО. Уникальные значения [M, B].



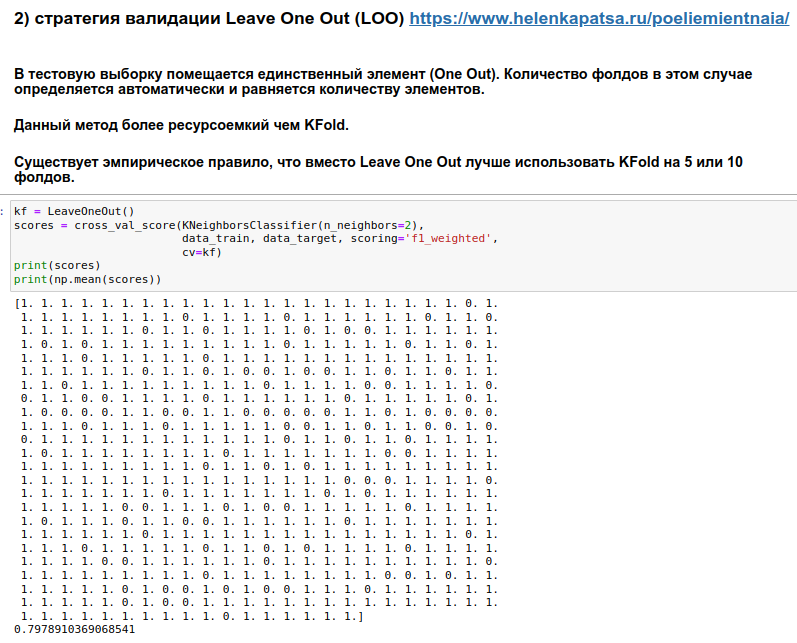


# **Стратегии валидации**

## 1) стратегия валидации k-fold

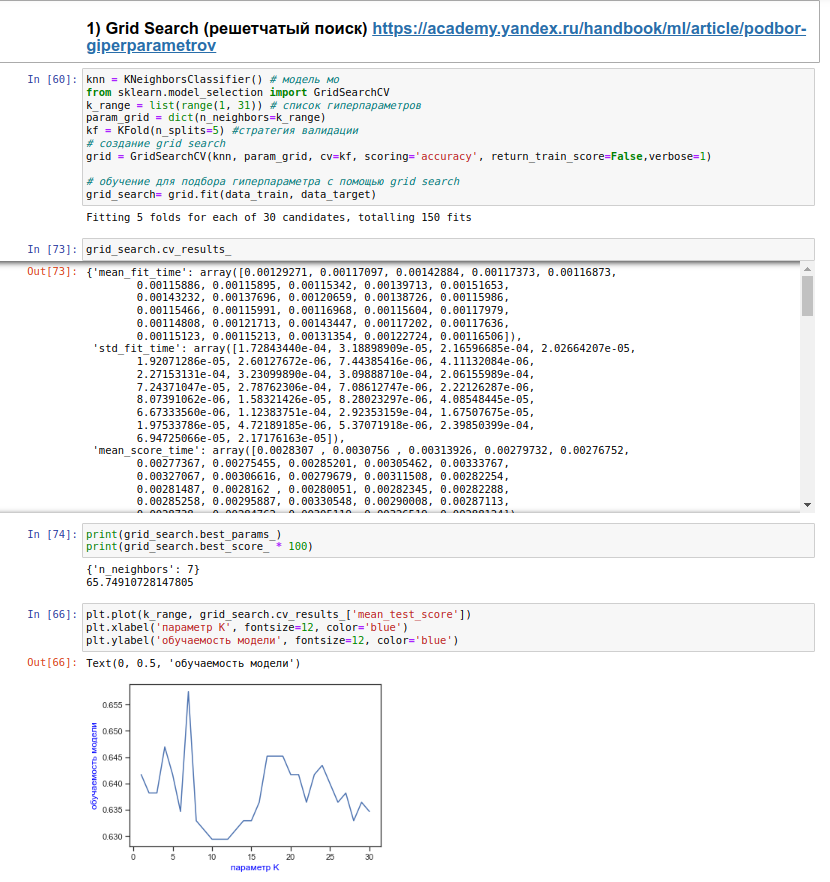


## 2) стратегия валидации Leave One Out (LOO)

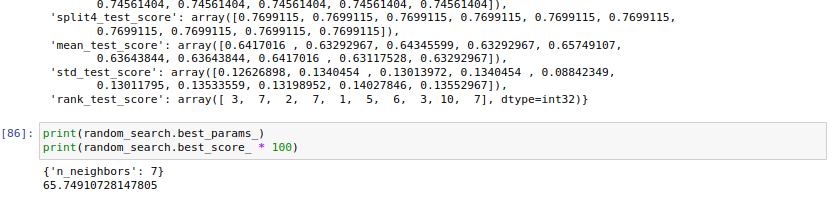
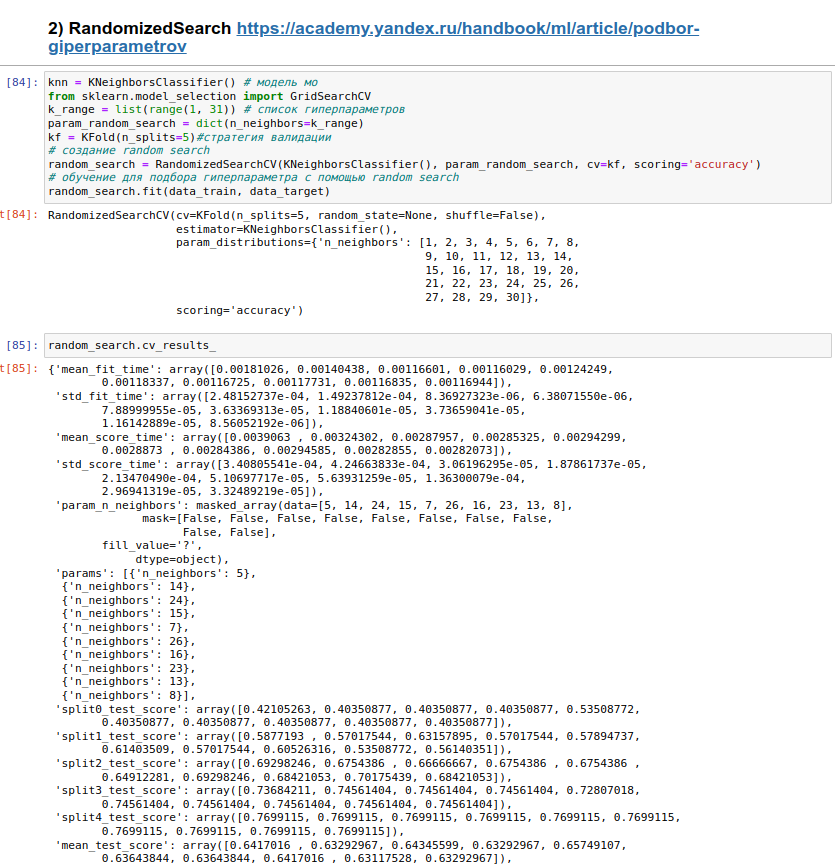


# **Подбор гиперпараметра K**

## 1) Grid Search (решетчатый поиск)



## 2) RandomizedSearch



# **Метрики качества классификации**



# **Вывод:**

Посмотрев на две матрицы, можно сделать вывод,что оптимальная можель намного лучше.

# **Вывод по лабораторной работе:**

# Познакомился с методами подбора гиперпараметра, с моделью k ближайших соседей. С крос-валидацией и ее стратегиями. А также с анализом моделей(созданная без анализа гиперпараметров и с подбором гиперпараметров). И стало понятно, что методы для подбора гиперпараметров имеют важную роль при создании модели машинного обучения