|  |
| --- |
|  |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования  **"МИРЭА - Российский технологический университет"**  **РТУ МИРЭА** |
| Институт информационных технологий (ИТ) |
| Кафедра прикладной математики (ПМ) |

|  |  |
| --- | --- |
| **ОТЧЕТ**  **ПО ПРАКТИЧЕСКОЙ РАБОТЕ №9** | |
| **по дисциплине** |  |
| **«Технологии и инструментарий анализа больших данных»** | |
| Выполнил студент группы ИКБО-20-19 | Московка А.А. |
| Принял семинарист кафедры Прикладной Математики | Десятников А.А. |

|  |  |  |
| --- | --- | --- |
| Практическая работы выполнены | « » 2022г. |  |
| «Зачтено» | « » 2022 г. |  |

Москва 2022

**Постановка задачи**

1. Найти данные для классификации. Данные в группе повторяться не должны! Предобработать данные, если это необходимо.
2. Изобразить гистограмму, которая показывает баланс классов. Сделать выводы.
3. Разбить выборку на тренировочную и тестовую. Тренировочная для обучения модели, тестовая для проверки ее качества.
4. Применить алгоритмы классификации: логистическая регрессия, SVM, KNN. Построить матрицу ошибок по результатам работы моделей (использовать confusion\_matrix из sklearn.metrics).
5. Сравнить результаты классификации, используя accuracy, precision, recall и f1-меру (можно использовать classification\_report из sklearn.metrics). Также сравнить время работы алгоритмов. Сделать выводы.
6. Оформить отчет о проделанной работе, написать выводы.

**Программный код с комментариями и результаты его выполнения**

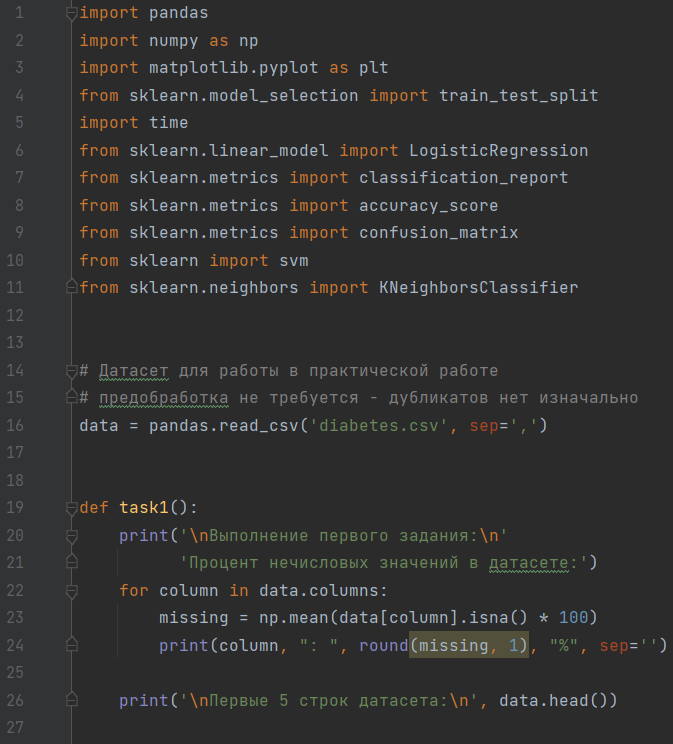


Рисунок 1 – Скриншот программного кода (Часть 1)



Рисунок 2 – Скриншот программного кода (Часть 2)



Рисунок 3 – Скриншот программного кода (Часть 3)

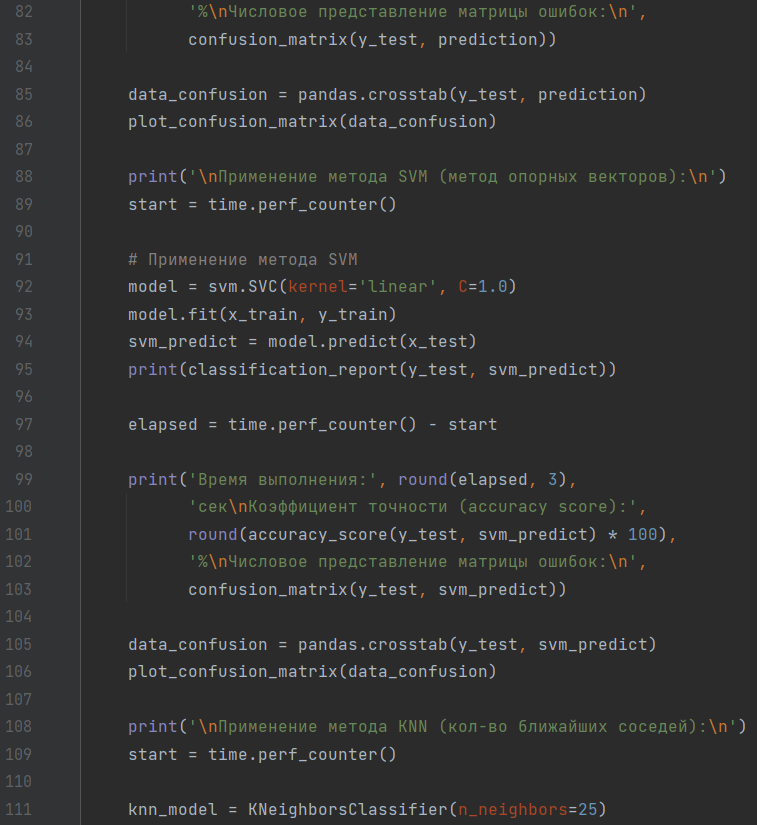


Рисунок 5 – Скриншот программного кода (Часть 4)

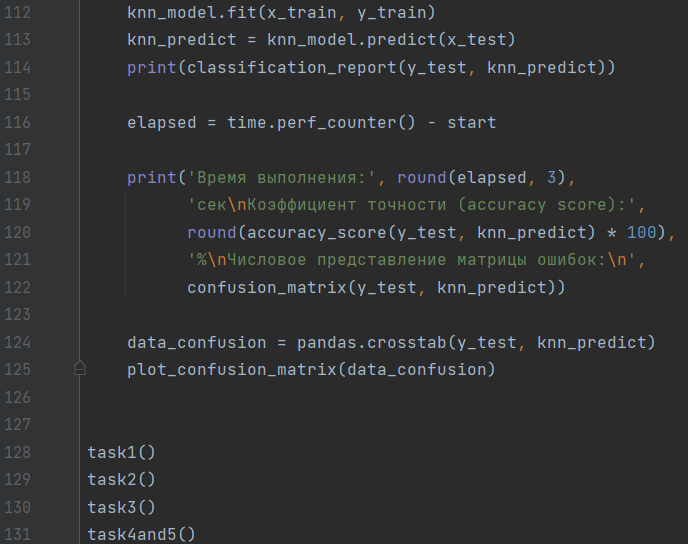


Рисунок 6 – Скриншот программного кода (Часть 5)

На следующих рисунках представлены скриншоты результатов выполнения программы со всеми заданиями с пояснениями, а также диаграммы (Рисунок 7-14).

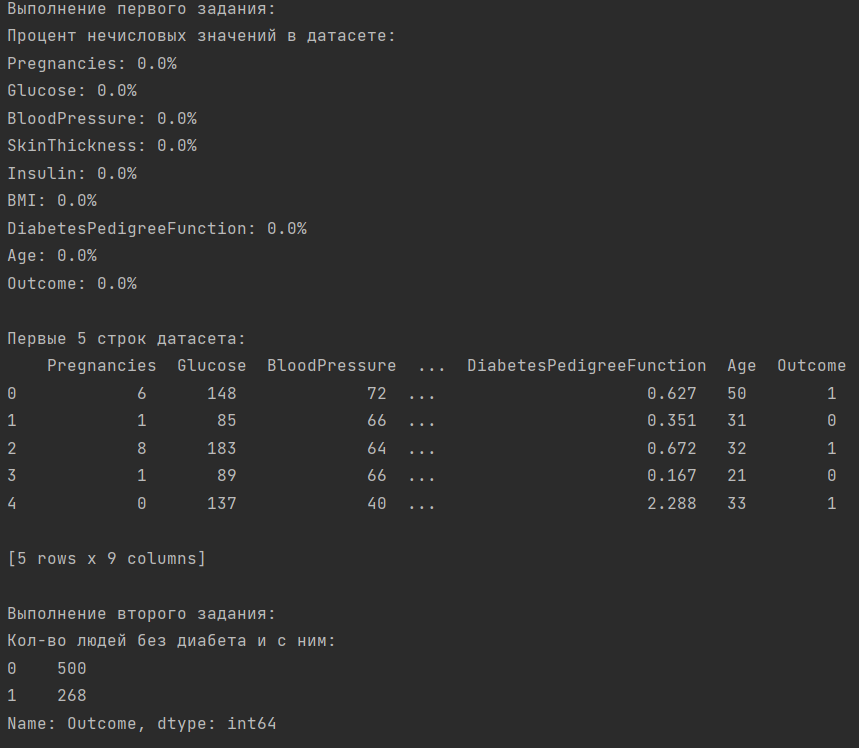


Рисунок 7 – Скриншот результата выполнения первого и второго заданий



Рисунок 8 – Скриншот результата выполнения второго задания

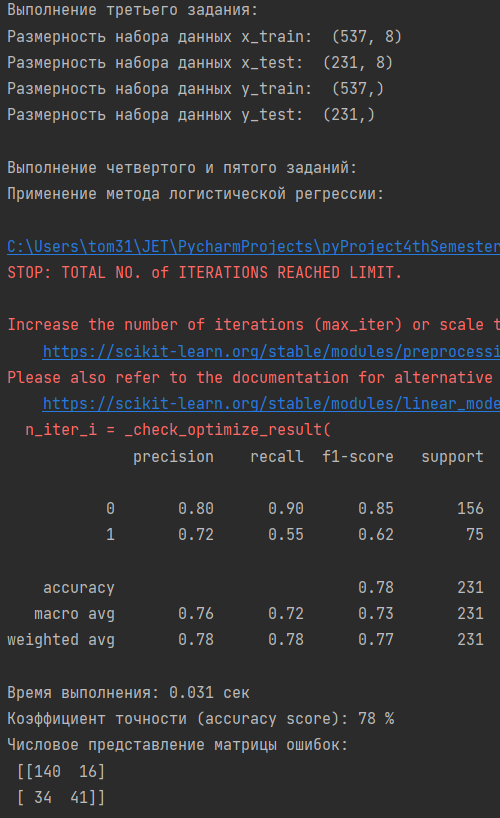


Рисунок 9 – Скриншот результата выполнения третьего, четвертого и пятого заданий

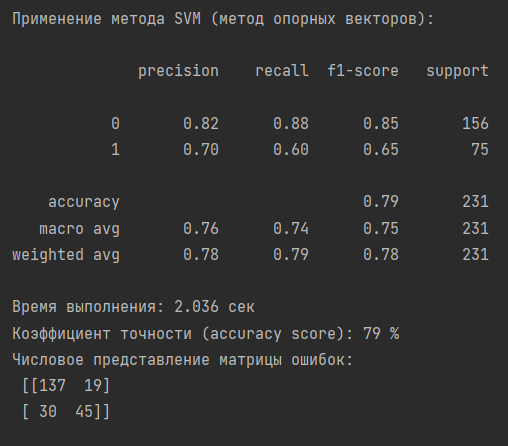


Рисунок 10 – Скриншот результата выполнения пятого задания (Часть 1)

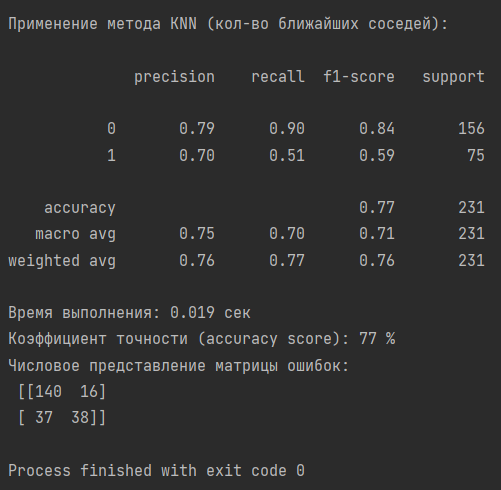


Рисунок 11 – Скриншот результата выполнения пятого задания (Часть 2)

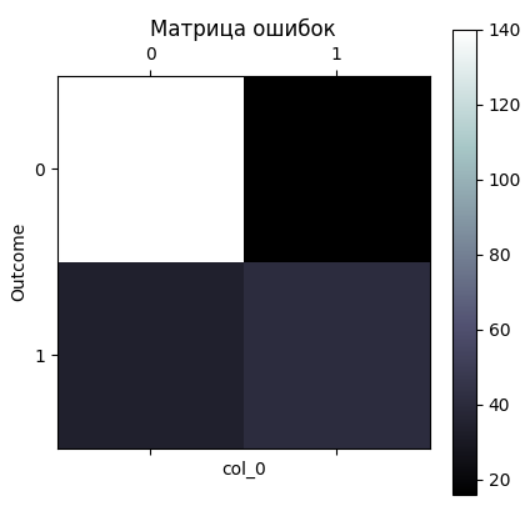


Рисунок 12 – Скриншот результата выполнения пятого задания (Матрица 1)

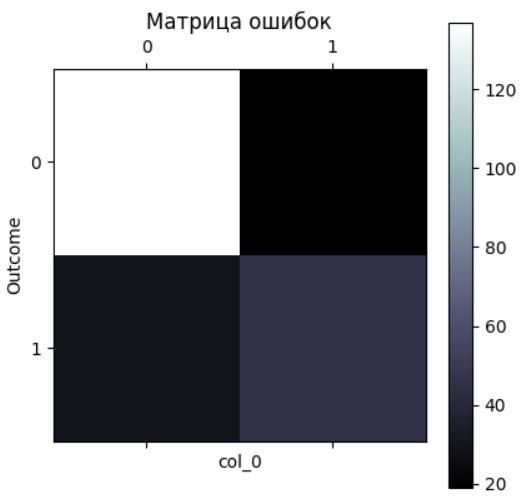


Рисунок 13 – Скриншот результата выполнения пятого задания (Матрица 2)

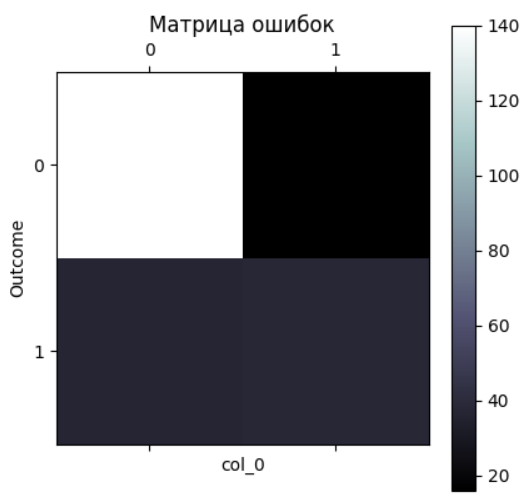


Рисунок 14 – Скриншот результата выполнения пятого задания (Матрица 3)

**Вывод**

В результате выполнения данной практической работы были найдены данные для классификации (данные по диабету), изображена гистограмма с отображением баланса классов, сделаны выводы по ее содержанию, разбита выборка на тренировочную и тестовую, применены и сравнены алгоритмы классификации логистическая регрессия, SVM, KNN, построена матрица ошибок по результатом работы моделей, сравнены результаты используя параметры accuracy, precision, recall и f1-меру, сравнено время работы алгоритмов, наконец, оформлен отчет о проделанной работе.