|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  | | МИНОБРНАУКИ РОССИИ | | | | Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА** – **Российский технологический университет»**  **РТУ МИРЭА** | | | | |
| Институт информационных технологий (ИТ) | |
| Кафедра прикладной математики | |

|  |  |
| --- | --- |
| **ОТЧЁТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ № 5** | |
| **по дисциплине «Технологии и инструментарий анализа больших данных»** | |
|  | |
| Выполнил студент группы ИКБО-20-21 | Фомичев Р.А. |
| Проверил ассистент кафедры ПМ ИИТ | Тетерин Н.Н. |

# Задание № 1

Найти данные для классификации. Данные в группе повторяться не должны. Предобработать данные, если это необходимо.

Код программы представлен на рисунке 1.



Рисунок 1 – Код программы

Вывод программы представлен на рисунке 2.

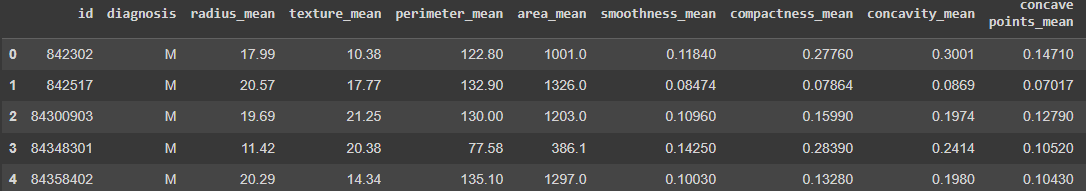


Рисунок 2 – Вывод программы

# Задание № 2

Изобразить гистограмму, которая показывает баланс классов. Сделать выводы.

Код программы представлен на рисунке 3.

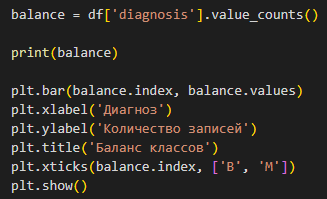


Рисунок 3 – Код программы

Вывод программы представлен на рисунке 4.

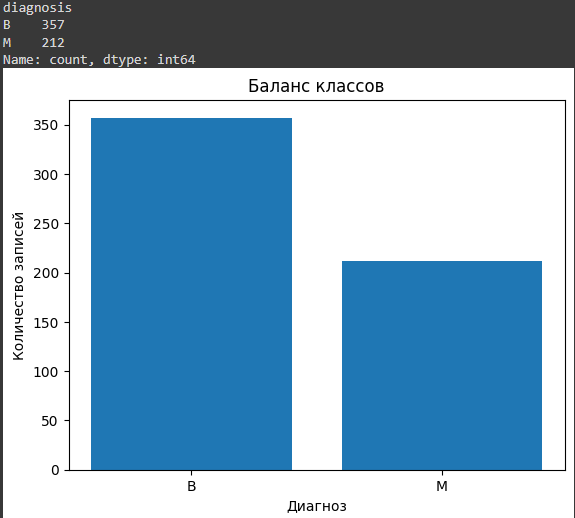


Рисунок 4 – Вывод программы

# Задание № 3

Разбить выборку на тренировочную и тестовую. Тренировочная для обучения модели, тестовая для проверки ее качества.

Код программы представлен на рисунке 5.

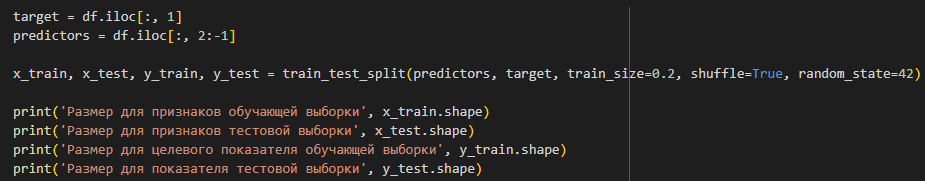


Рисунок 5 – Код программы

Вывод программы представлен на рисунке 6.

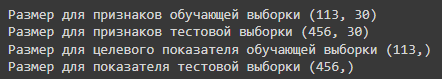


Рисунок 6 – Вывод программы

# Задание № 4

Применить алгоритмы классификации: логистическая регрессия, SVM, KNN. Построить матрицу ошибок по результатам работы моделей (использовать confusion\_matrix из sklearn.metrics)

Код программы представлен на рисунке 7.

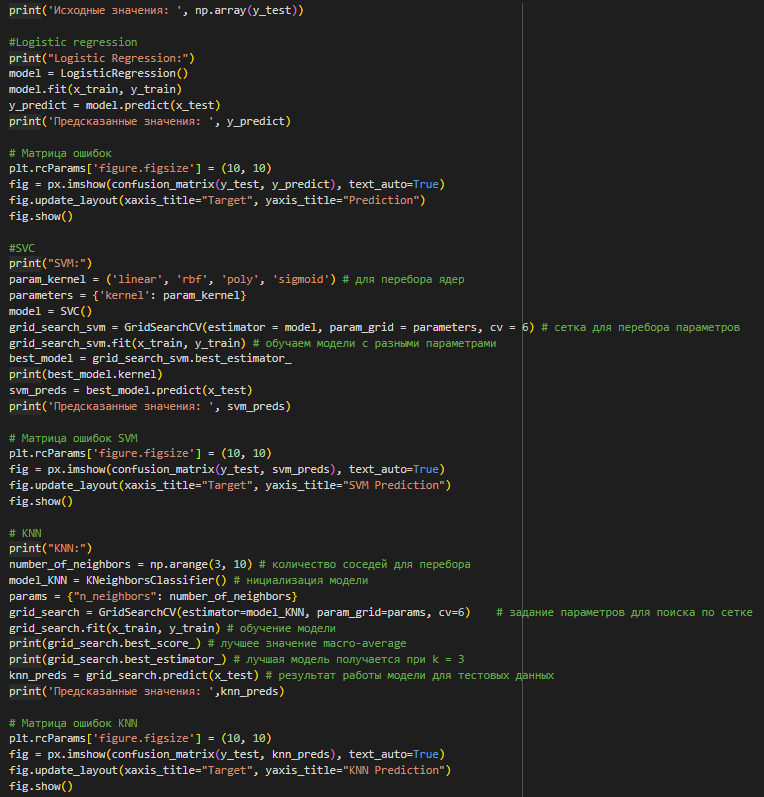


Рисунок 7 – Код программы

Вывод программы представлен на рисунках 8 – 10.

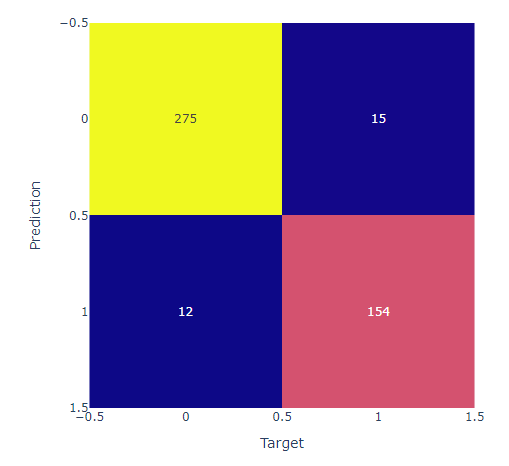


Рисунок 8 –Logistic Regression

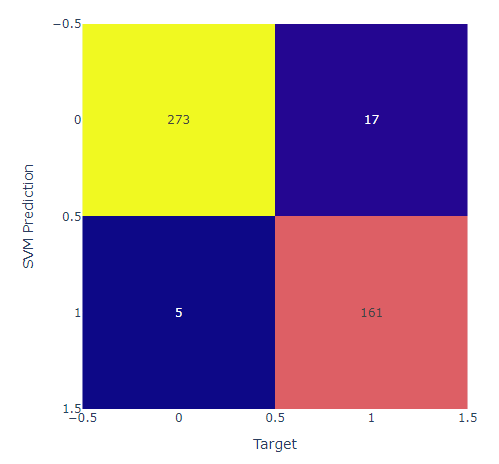


Рисунок 9 – SVM

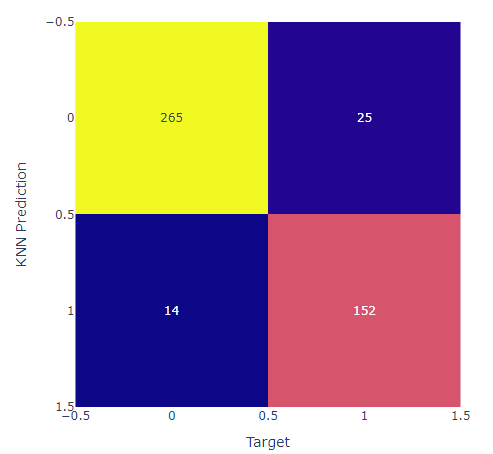
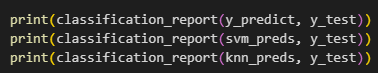


Рисунок 10 – KNN

# Задание № 5

Сравнить результаты классификации, используя accuracy, precision, recall и f1-меру (можно использовать classification\_report из sklearn.metrics). Сделать выводы.

Код представлен на рисунке 11.



Вывод программы представлен на рисунке 12.

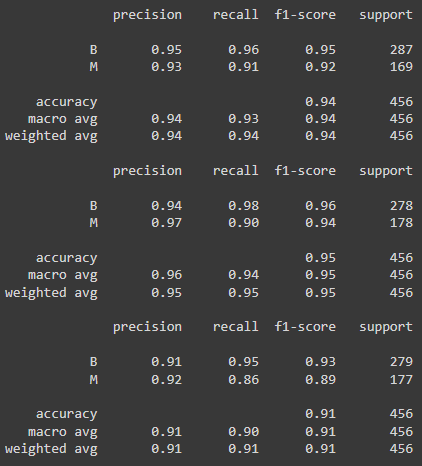


Рисунок 12 – Вывод программы