**Московский государственный технический   
университет им. Н. Э. Баумана**

Курс «Технологии машинного обучения»

Отчёт по рубежному контролю №1

«Технологии разведочного анализа и обработки данных.»

Вариант № 17

Выполнил Проверил:  
Сайфутдинов Р.И. Гапанюк Ю.Е.  
группа ИУ5-62Б

Дата: 15.04.25 Дата:

Подпись: Подпись:

Москва, 2025 г.

**Задание:**

Номер варианта: **17**

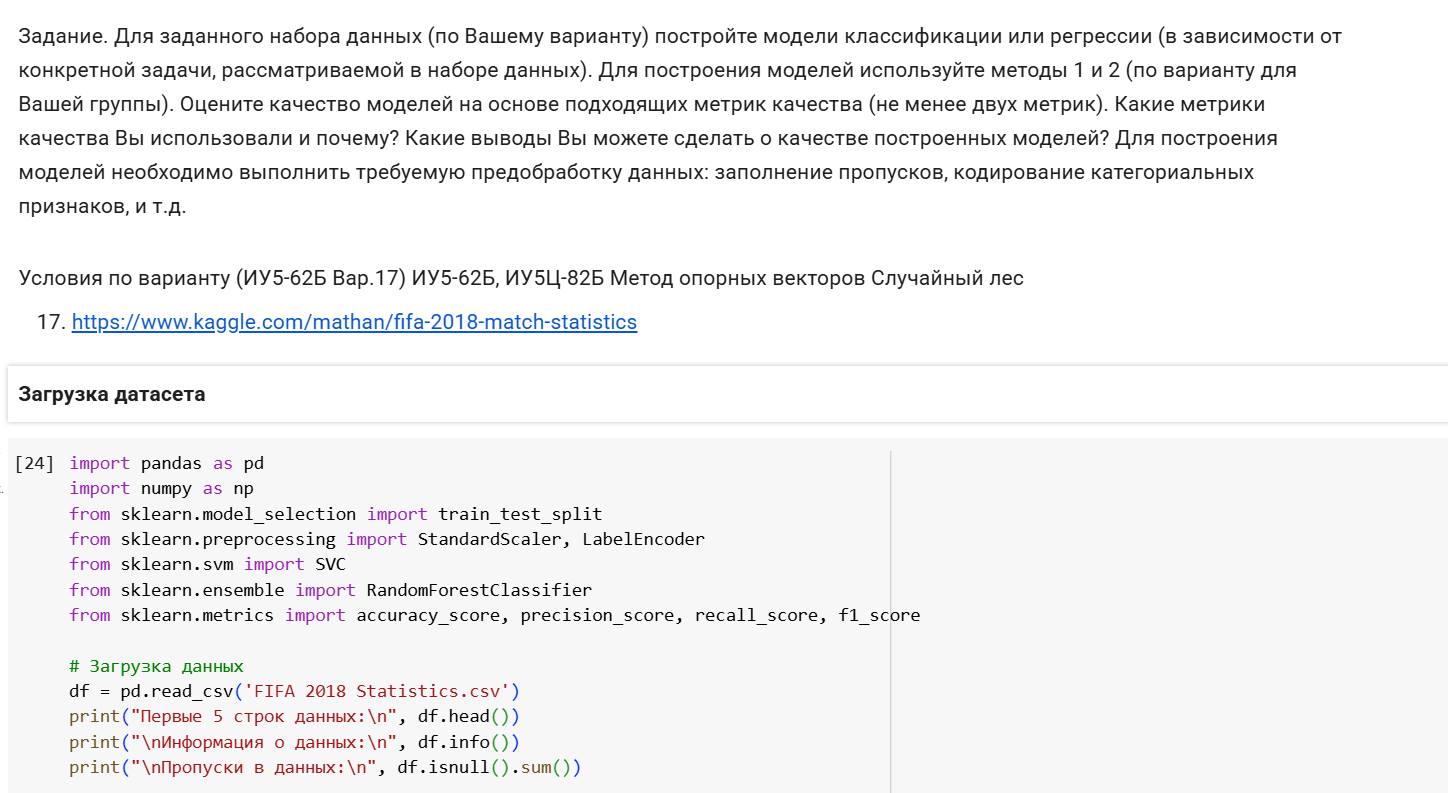
<https://www.kaggle.com/mathan/fifa-2018-match-statistics>

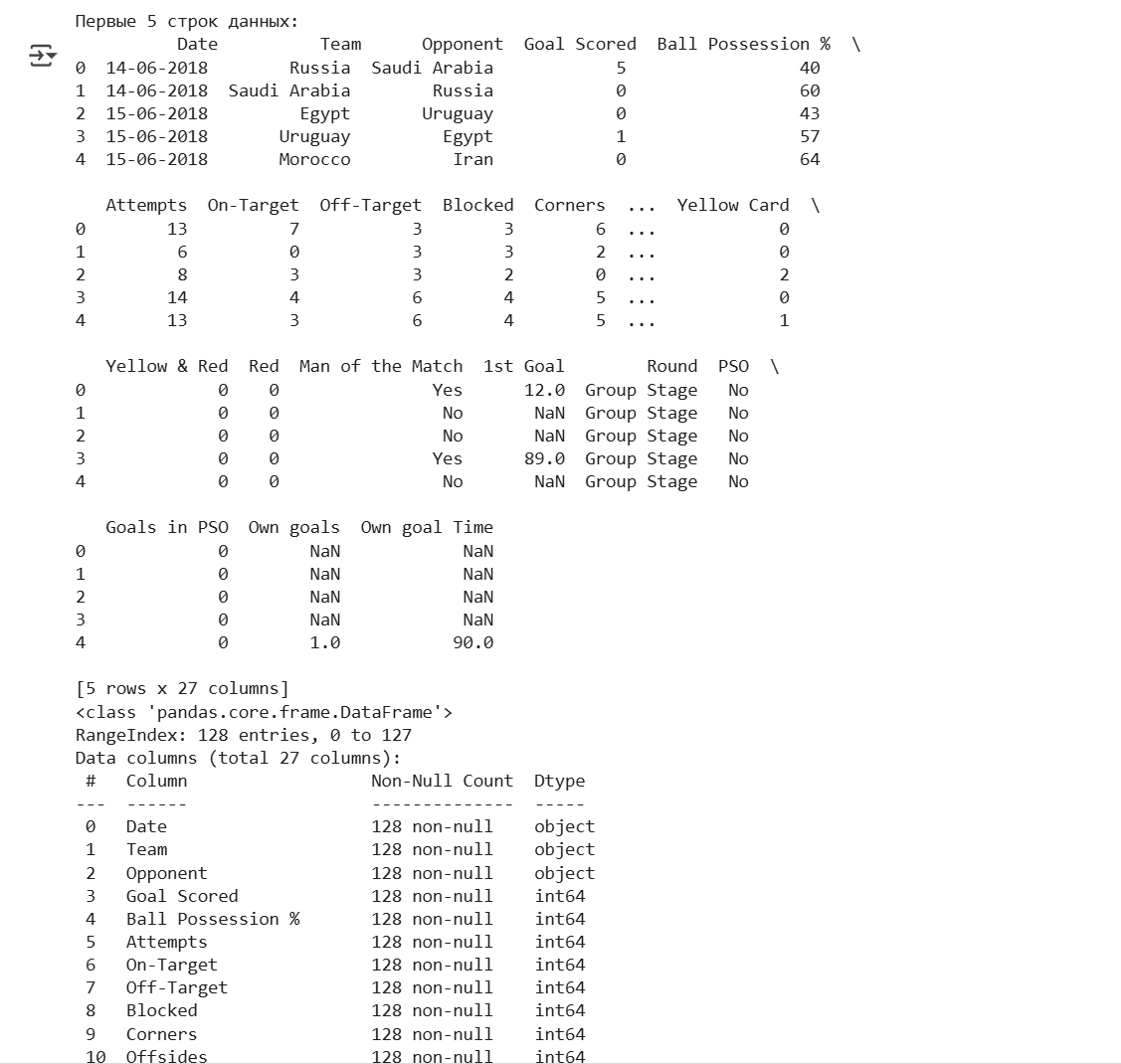
|  |  |  |
| --- | --- | --- |
| Группа | Метод №1 | Метод №2 |
| ИУ5-62Б, ИУ5Ц-82Б | Метод опорных векторов | Случайный лес |

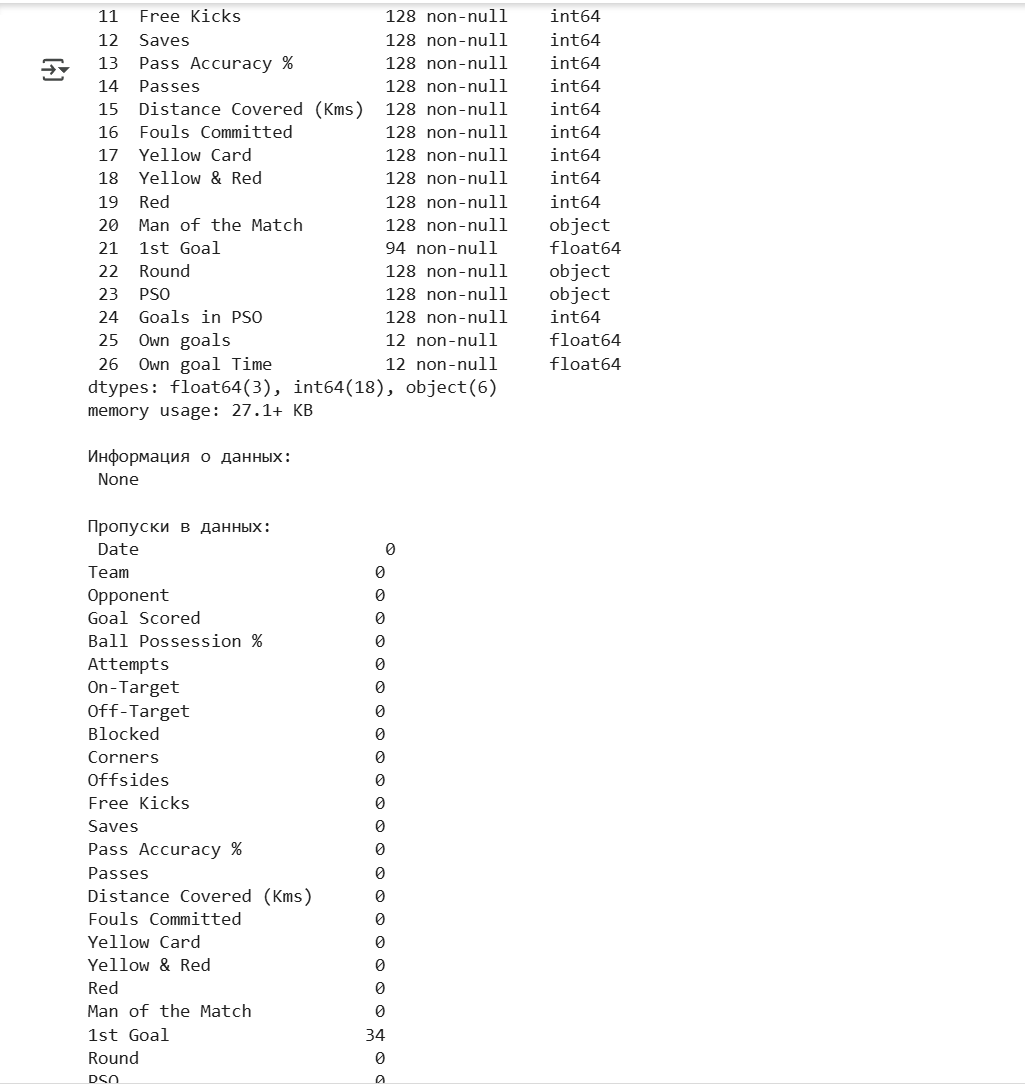
**Задача №3.**

Для заданного набора данных (по Вашему варианту) постройте модели классификации или регрессии (в зависимости от конкретной задачи, рассматриваемой в наборе данных). Для построения моделей используйте методы 1 и 2 (по варианту для Вашей группы). Оцените качество моделей на основе подходящих метрик качества (не менее двух метрик). Какие метрики качества Вы использовали и почему? Какие выводы Вы можете сделать о качестве построенных моделей? Для построения моделей необходимо выполнить требуемую предобработку данных: заполнение пропусков, кодирование категориальных признаков, и т.д.

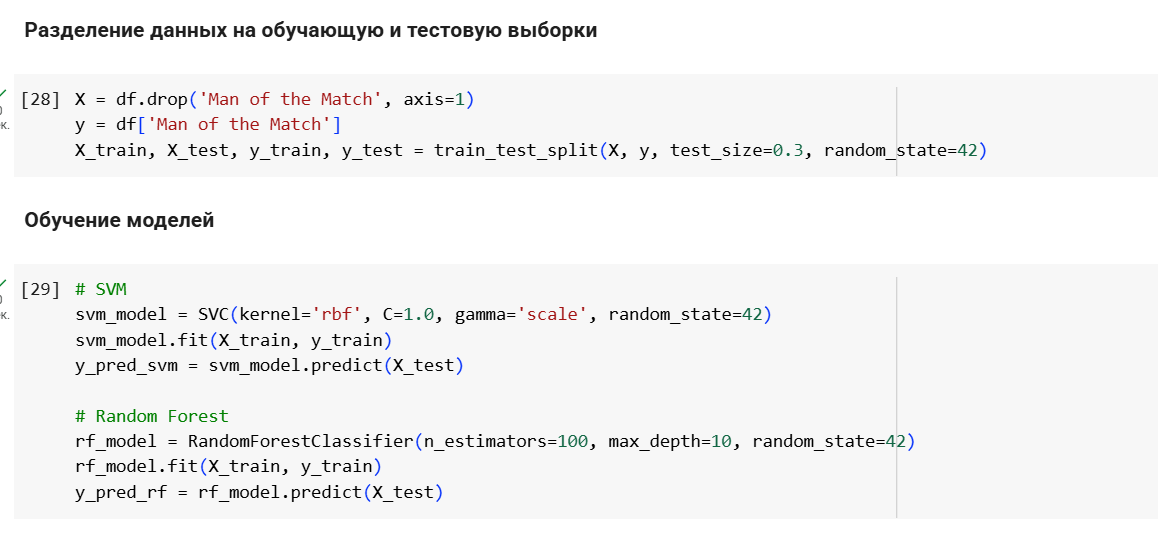
**Ход выполнения:**













**Какие метрики использовались и почему?**

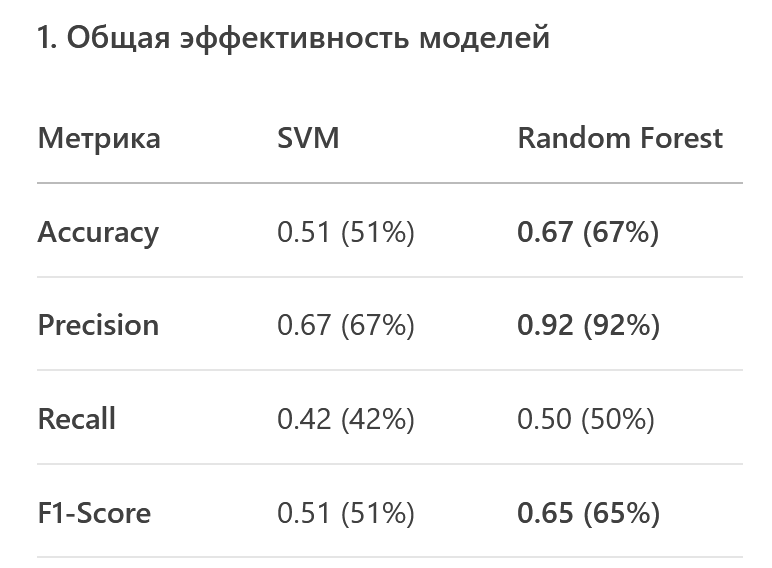
**Accuracy** — общая точность классификации.

**Precision** — доля верно предсказанных Man of the Match среди всех предсказанных.

**Recall** — доль верно предсказанных Man of the Match среди всех реальных.

**F1-Score** — гармоническое среднее Precision и Recall (важно при дисбалансе классов).

**Сравнение моделей**



**Random Forest показал лучшие результаты** по всем метрикам.

SVM хуже справляется с дисбалансом классов (меньше Recall).

Random Forest лучше обобщает данные благодаря ансамблевому подходу.