Задание №4

В этом задании вам предстоит классифицировать типы соевых бобов (http://archive.ics.uci.edu/ml/datasets/Soybean+%28Large%29).
В рамках этого задания вам необходимо сделать следующее:
1. Сделать EDA для ваших данных (5 баллов):
   – проанализируйте какие переменные есть, возможно стоит построить графики
   – проверить наличие NA, что с ними сделать – на ваше усмотрение
   – проверить наличие скоррелированных переменных в ваших данных, визуализируйте это, за вами выбор оставить эти переменные или нет.
2. Разделите ваш датасет на тестовую и тренировочную выборки в таком отношении, что бы все классы были представлены в обоих выборках и были стратифицированны (равное распределение классов, нужно указать доп. опцию в фунции `train_test_split`) (2 балла).
3. Обучите ваш классификатор (предлагаю воспользоваться методом RandomForestClassifier) на вашей обучающей выборке (2 балла).
4. В случае тестовой выборки удалите информацию о принадлежности к классам (1 балл).
5. Предскажите значения классов используя ваш классификатор (2 балла).
6. Оцените качество вашего классификатора используя известные вам метрики (F-мера и матрица неточностей). Какие выводы вы можете сделать (5 баллов).
7. Выведите топ-3 признаков, которые оказались самыми важными при классификации (2 балла).
8. С помощью GridSearch оптимизируйте ваш классификатор и сравните его с полученным изначально (4 баллов).