Была применена логистическая регрессия – алгоритм классификации, который применяется для прогнозирования вероятности зависимой переменной.

После загрузки и просмотра датасета, были удалены столбцы 1 (по условию) и 7 (в нём практически одни Nan).  
Далее в столбце 6 были заменены значения Yes и No на 1 и 0.  
После этого выводится соотношение классов в столбце PD. Классы несбалансированные (0 – 79 %, 1 – 21 %).

Далее происходит визуализация некоторых признаков, чтобы посмотреть, какие показатели могут служить хорошим фактором просмотра итоговой переменной. Например, столбец 10 не является хорошим прогнозирующим фактором.

Далее увеличивается частота выборки 1 из PD при помощи Over-Sampling. Это синтетическая техника дублирования примеров мониторинга класса. Создаются синтетические образцы на основе выборок минорного класса, а не просто берутся их копии.  
Over-Sampling я применяю только на обучающих данных, чтобы информация из тестовых данных не использовалась для создания синтетических сведений и не произошло утечки в обучающую модель.   
После этого соотношение классов в столбце PD равное (по 50 %).

После этого для отбора признаков я применяю рекурсивное устранение признаков (RFE), матрицу корреляции, statsmodel.

RFE основывается на повторяющемся конструировании модели и выборке признака, который выполняется хуже всех. Затем происходит отделение этого признака и повторения цикла с оставшимися.

Проведя анализ, было решено для построения модели оставить следующие признаки: [8, 9, 11, 14, 19, 21, 22].

Далее осуществляется прогнозирование результатов и вычисление точности.  
Выборка делится на обучающую и тестовую в соотношении 9 к 1.  
Точность классификатора логистической регрессии для тестового набора составила 78 %.  
Как было сказано в условии, не были применены для классификации признаки 17 и 18. При их использовании точность классификатора составляет 99 %.  
Также, классификатор, возможно, был бы точнее, если бы обучающая выборка была больше. Логистическая регрессия требует больших размеров выборки.

Далее выводятся метрики валидации результата. Precision, recall, F1, ROC AUC.