***Отчет по Task3***

**Первый этап** — подготовка данных. Преобразование данных из .txt в .csv точно такое же, как и во втором задании. Далее возпользуемся ProfileReport, чтобы получить детальную информацию по датасету. Можно видеть, что во всех столбцах есть пропуски, и во многих из них нули составляют очень большу часть от всего столбца — от 60% до 98%. Однако есть и столбцы, в которых пропущено всего 3-9% данных. Учитывая эти факты, а также то, все 128 предикторов — непрерывные величины, а не категориальные, то удаление строк или столбцов с нулями приведет к потере почти всех данных, поэтому можно попробовать следующий алгоритм: на начальном этапе оставить только те столбцы, где кол-во пропущенных данных не превышает 10%, и заполнить эти пропуски (средним, методом соседей, линейной регрессией или еще каких образом); на следующем этапе добавлять столбцы, где пропущенных данных больше и заполнять пропуски теми же методами; на последнем же этап добавить столбцы, в которых пропусков больше 90%, и вместо нулей вставить средние значение по соотвествующим столбцам. Изучив отчет по данному датасету, оказалось, что в каждом столбце есть нули. Все столбцы разбиты на 5 групп: <10%, 10-30%, 30-60%, 60-90%, >90%, по кол-ву в них нулей. Второй этап — заполнение пропусков. Пропуски в первой группе будем заполнять средним значением соотвествующего столбца. Таких столбцов мало, и число пропусков мало, по сравнению с другими, поэтому после заполнения эти столбы вполне сгодятся в качестве базы. В пятой группе пропуски также будем заполнять средними значениями, так как их слишком много, а заполнение их, основываясь на заполнении столбцов из други групп, может повлечь за собой накопление излишнего шума и ошибки. При работе со второй группой каждый столбец из нее будет прогнозируемой величиной, а столбцы из первой группы — предикторами. Соотвественно номера строк в столбцах из второй группы, не содержащие нулей, будут обучающими. Будем последовательно обучать модель на каждой паре «столбцы первой группы»-«столбец из второй группы». На этом этапе в качестве моедели будем использовать MLPRegressor, которая хороша в аппроксимации нелинейных ф-ий.