В качестве испытания представлен набор данных Mechanical Data Set. В архиве находятся:

1. Файл с набором данных.
2. Файл с записями из набора данных, не включенными в основной набор (обычно - не используются).
3. Файл с описанием набора данных (английский язык, можно воспользоваться любым удобным переводчиком).

Набор данных описывает результаты контроля состояния электро-механического оборудования (насосов) с применением методов вибрационной диагностики. В качестве выходного параметра рассматривается состояние, характеризующееся либо полной исправностью, либо одной или несколькими неисправностями (подробно – в файле с описанием). Таким образом – решается задача классификации.

Необходимо:

1. Ознакомиться с описанием набора данных, дать краткую интерпретацию смыслового содержания. Пояснить атрибуты (входные параметры) и возможные классы состояний диагностируемых объектов.
2. Ознакомиться с файлом с данными, дать краткую характеристику структуре представления данных в нем.
3. Сформировать (разработать) и кратко описать схему подготовки, приведения к необходимому формату и предобработки данных к анализу. Реализовать схему, пояснив шаги, выполняемые «вручную» и автоматизированным способом.
4. Провести статистический анализ исходных данных, определив наиболее значимые входные переменные и/или сформировав новые входные переменные, например, методом главных компонент (приветствуются и другие методы). Опционально: В случае сокращения размерности пространства входов провести анализ структуры данных, разделимости выборки, визуализировав ее в «сокращенном» пространстве.
5. Выполнив программную реализацию, сформировать не менее 5 различных классификаторов и провести статистическое исследованием методом 5-кратной перекрестной проверки. Привести результаты классификации объектов в выборках, а также - примеры результатов ROC-AUC анализа и матрицы ошибок для выбранных типов классификаторов с пояснениями. Опционально: осуществить также выбор гиперпараметров методов, используемых для построения классификаторов выбранного типа.
6. Выполнить проектирование и реализовать простой графический интерфейс приложения для выполнения и указания параметров функций, реализованных программно.
7. Привести код разработанных программ с пояснением.
8. Сформировать отчет о проведенных исследованиях предлагаемого набора данных, включив в него, в том числе, графические и табличные формы представления результатов.