

Project Idea #2

Data Mining Project on Diabetes Prediction

Diabetes is one of the most common and hazardous diseases on the planet. It requires a lot of care and proper medication to keep the disease in control. If you are curious about data mining projects in healthcare, you should explore the diabetes dataset.

Dataset: The data consist of medical information, laboratory analysis... etc. The data that have been entered initially into the system are: No. of Patient, Sugar Level Blood, Age, Gender, Creatinine ratio(Cr), Body Mass Index (BMI), Urea, Cholesterol (Chol), Fasting lipid profile, including total, LDL, VLDL, Triglycerides(TG) and HDL Cholesterol , HBA1C, Class (the patient's diabetes disease class may be Diabetic, Non-Diabetic, or Predict-Diabetic).

The project implementation steps will go as follows:

- Understand the dataset attributes
- Apply the required data cleaning methods
- Detect the outliers
- Implement different classification models to investigate the performance of each classifier on diabetes datasets.
- Mention your observations and study the parameters (features) to determine the major factors affecting the onset of diabetes
- (What percentage of younger people are prone to be diagnosed with diabetes disease?
Are women more prone to diabetes, or is it the other way? etc.)
- Visualize the result of data in plots (discover the potential plots to describe result)