In this project, you must detect diabetes people. There is a dataset (link). This dataset has 8 features in 2 classes. The Bayesian analysis must be Implemented for this dataset.

- **Data vitalization and Preprocessing:**
  - Visualize the relationship between variables using scatter plot (According to Figure 1 in the appendix).
  - Split dataset for Train and Test, 70% and 30% of data respectively.
- **Modeling:**
  - Implement Gaussian Naïve Bayes classifier by Train dataset.
- **Evaluation:**
  - Evaluate the model by Train and Test dataset.
  - Evaluate the model based on 4-fold cross validation.
  - Polt two confusion matrix for the classifier by Train and Test dataset.
- **Model efficiency improvement (Extra credit):**
  - Select the best four features according to the Data vitalization part then make a new dataset.
  - Say, why these features are the best.
  - Retrain the model with the new dataset.
  - Evaluate the new model with the new dataset.
- **Apply thresholding method (Extra credit):**
  - Add a parameter to the model for adjusting the classifier.
  - Set this parameter for detecting diabetes when the probability of diabetes is more than 40%.

Appendix:

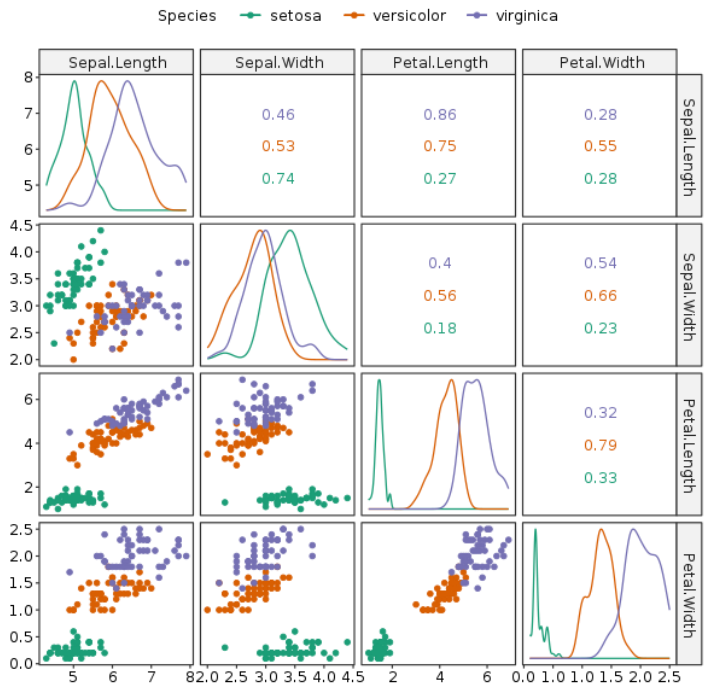

Figure 1: Scotter plots and line charts are used in descriptive statistics to show the relationship between Iris dataset features.