# Ensemble Techniques

1. **Business Problem**
   1. **What is the business objective?**
   2. **Are there any constraints?**
2. **Work on each feature of the dataset to create a data dictionary as displayed in the below image:**



**2.1. Make a table as shown above and provide information about the features such as its data type and its relevance to the model building. And if not relevant, provide reasons and a description of the feature.**

**Using R and Python codes perform:**

1. **Data Pre-processing**

**3.1 Data Cleaning, Feature Engineering, etc.**

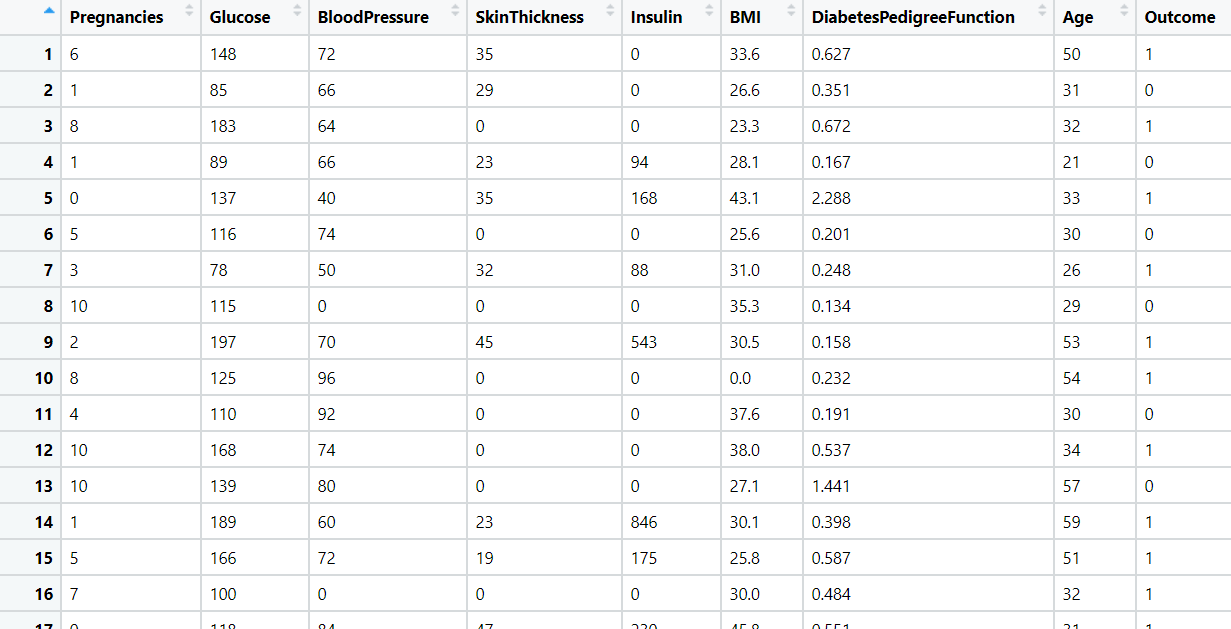
**3.2 Outlier Treatment.**

1. **Exploratory Data Analysis (EDA):**
   1. **Summary.**
   2. **Univariate analysis.**
   3. **Bivariate analysis.**
2. **Model Building**
   1. **Build the model on the scaled data (try multiple options).**
   2. **Perform Bagging, Boosting, Voting, Stacking on given datasets.**
   3. **Train and Test the data, use grid search cross validation, compare accuracies using confusion matrix.**
   4. **Briefly explain the model output in the documentation.**
3. **Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.**
4. **Model Building**
   1. **Build the model on the scaled data (try multiple options).**
   2. **Perform Bagging Boosting (adaboost, fastadaboost, Xgboost), Stacking, Voting on the given datasets in Hands on Material.**
   3. **Train and Test the model and compare accuracies by building a confusion matrix and use different hyperparameters. Also use GridSearchCV to improve your model performance.**
   4. **Briefly explain the model output in the documentation.**

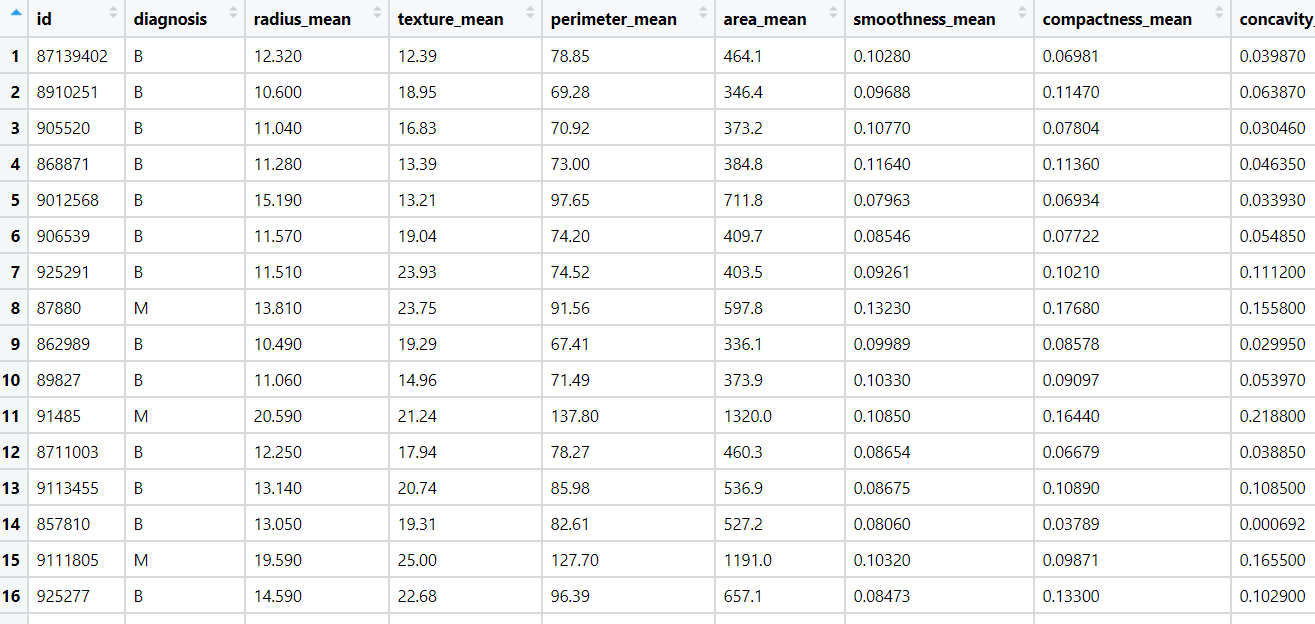
1. **Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?**

**Problem Statements:**

1. Given is the diabetes dataset. Build an ensemble model to correctly classify the outcome variable and improve your model prediction by using GridSearchCV. You must apply Bagging, Boosting, Stacking, and Voting on the dataset.



1. Most cancers form a lump called a tumour. But not all lumps are cancerous. Doctors extract a sample from the lump and examine it to find out if it’s cancer or not. Lumps that are not cancerous are called benign (be-NINE). Lumps that are cancerous are called malignant (muh-LIG-nunt). Obtaining incorrect results (false positives and false negatives) especially in a medical condition such as cancer is dangerous. So, perform Bagging, Boosting, Stacking, and Voting algorithms to increase model performance and provide your insights in the documentation.



1. A sample of global companies and their ratings are given for the cocoa bean production along with the location of the beans being used. Identify the important features in the analysis and accurately classify the companies based on their ratings and draw insights from the data. Build ensemble models such as Bagging, Boosting, Stacking, and Voting on the dataset given.

**A screenshot of a computer

Description automatically generated**

1. Data privacy is always an important factor to safeguard their customers' details. For this, password strength is an important metric to track. Build an ensemble model to classify the user’s password strength.

A screenshot of a cell phone

Description automatically generated