Report For Few Features:

1. The data has several Features for patient who either are normal or hypertensive/hypotensive patients.
2. Initially the following features where selected.
   1. Level\_of\_Hemoglobin,Genetic\_Pedigree\_Coefficient as features
   2. Blood\_Pressure\_Abnormality as classification features
3. Fills the Nans and Null in the numerical features with mean value.
4. Removes the data within 3 standard deviation using normal distribution
5. Normalizes the data to same scale
6. Creates an imputer to fill Nan for categorical variable
7. Split data into train and test with 70% train and 30% test
8. Reshapes the data
9. Create Three Ensemble Classification Model
   1. Random forest
   2. Gradient boosted
   3. Adaboost
10. For Each Classification model
    1. Creates a grid to search for hyper parameters for the best accuracy model using crossvalidation with 5 folds
    2. Fit the train and test data
    3. Store Accuracy,Roc Auc score and the confusion matrix
11. creates voting classifier using the above 3 classification model
    1. Fit the train and test data
    2. Store the Accuracy,Roc Auc score and the confusion matrix
12. Finally the best model had the following scores
    1. **Name:Gradient Boosting**
    2. **Accuracy: 0.896667**
    3. **Roc\_auc: 0.897093**
    4. **Confusion Matrix**

****

Report For All The Features:

1. The data has several Features for patient who either are normal or hypertensive/hypotensive patients.
2. Initially the following features where selected.
   1. All the features where included
   2. Blood\_Pressure\_Abnormality as classification features
3. Fills the Nans and Null in the numerical features with mean value.
4. Removes the data within 3 standard deviation using normal distribution
5. Normalizes the data to same scale
6. Creates an imputer to fill Nan for categorical variable
7. Split data into train and test with 70% train and 30% test
8. Reshapes the data
9. Create Three Ensemble Classification Model
   1. Random forest
   2. Gradient boosted
   3. Adaboost
10. For Each Classification model
    1. Creates a grid to search for hyper parameters for the best accuracy model using crossvalidation with 5 folds
    2. Fit the train and test data
    3. Store Accuracy,Roc Auc score and the confusion matrix
11. creates voting classifier using the above 3 classification model
    1. Fit the train and test data
    2. Store the Accuracy,Roc Auc score and the confusion matrix
12. Finally the best model had the following scores
    1. **Name:Random Forest**
    2. **Accuracy: 0.933333**
    3. **Roc\_auc: 0.933204**
    4. **Confusion Matrix**

****

**Conclusion:**

**This Clearly show that when features are added the accuracy of the model increases from 89 to 93 percentage**

**Code Link:**

**https://github.com/akash731/Absolute\_data\_assignement/blob/master/Absolute\_data\_Assignment.ipynb**