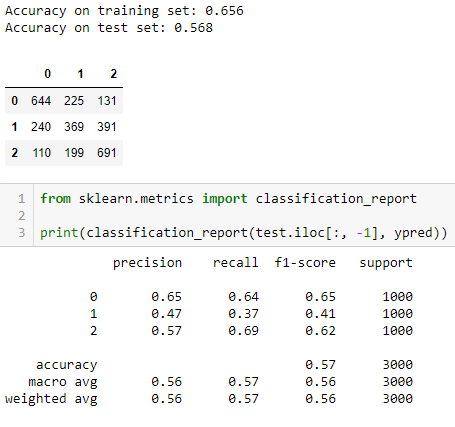
**Summary of hypertension prediction experiments**

1. **Hypertension classification**

The first experiment is hypertension classification without considering SBP and DBP values. The features used for this experiment includes both numerical and categorical values

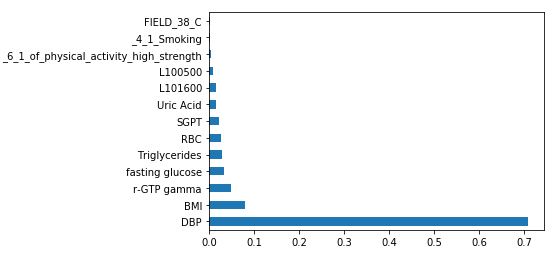
* + 1. BMI
    2. r-GTP gamma
    3. fasting glucose
    4. Triglycerides
    5. AGE
    6. SGPT
    7. SEX
    8. L101400 – LDH, legacy dehydratase
    9. Uric Acid
    10. L101600 - Alkaline phosphatase
    11. L100500 - Creatinine
    12. 6\_1\_of\_physical\_activity\_high\_strength \_
    13. 4\_1\_Smoking
    14. FIELD\_38\_C - Drinking\_5\_1\_days

The overall classification accuracy is 56.8%. which is quite low. There are several factors that affect the blood pressure value throughout the day other than the features included here, such as stress level, anger, salty food consumption, anxiety levels, and others. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8316330>

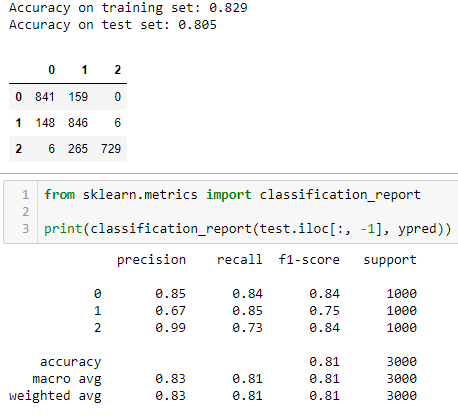


1. **Classification of SBP values**
   1. **Using DBP values**

In this section, SBP values are classified into 3 classes using the features which are highly correlated to SBP, including DBP.



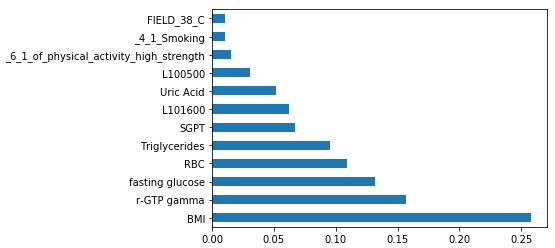
As we can see from the chart the DBP and SBP values are highly correlated. Consequently, the classification result of the SBP values is high. Not only the high correlation between the values result in high accuracy but also the fact that the class boundary is also defined by the DBP values. However, if we exclude the DBP values the classification accuracy suffers.



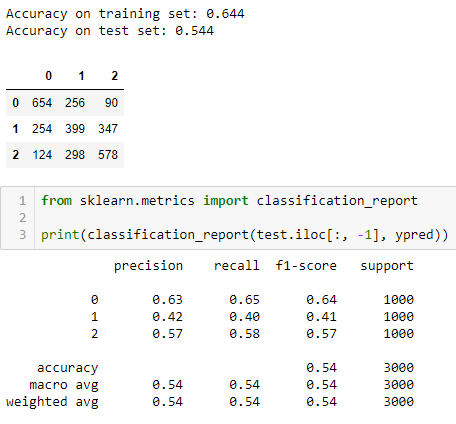
* 1. **Without using the DBP values**

As mentioned above, when we excluded DBP from the features list, the accuracy of the classification is affected negatively. The list of the features used is shown below. The best classification accuracy for this feature set is 54.4%

* + BMI
  + RBC
  + Uric Acid
  + Triglycerides
  + Fasting glucose
  + L100500 - Creatinine
  + r-GTP gamma
  + SGPT
  + L101600 - Alkaline phosphatase
  + 4\_1\_Smoking
  + FIELD\_38\_C - Drinking\_5\_1\_days
  + 6\_1\_of\_physical\_activity\_high\_strength



Random Forest Classifier



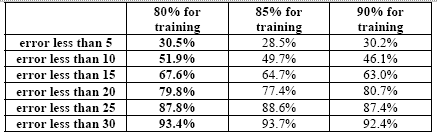
1. **SBP regression experiments**

In this section, regression experimental results are presented. The experiments have three sections

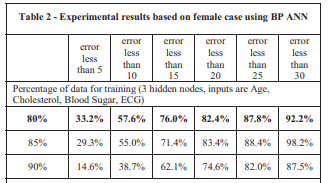
* Regression experiment including DBP as a feature
* Regression experiment excluding DBP as a feature
* Regression experiment for predicting next year SBP values

In all of the experiments the r2 square value lower than 70%, even if we include DBP values as a feature for the prediction model. However, the following research papers used the absolute difference as a metric to evaluate their prediction models.

1. [Predicting Systolic Blood Pressure Using Machine Learning](https://ieeexplore.ieee.org/abstract/document/7069529)



1. [Bio-medical Application on Predicting Systolic Blood](https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7184916)



The results from the first experiment (Regression experiment including DBP as a feature) show that the predictor has predicted 50% of the time with absolute difference of less than 5 percent. However, when we exclude DBP the metric is reduced to 35.4%. the next year's prediction model also has 42.2% for the same metric, which is better than the one presented in the papers listed above.

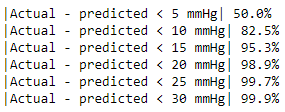
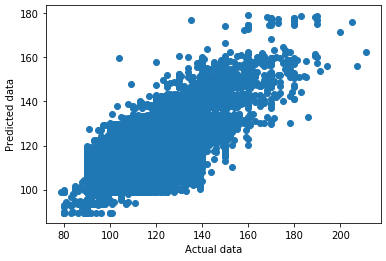
* 1. **Regression experiment including DBP as a feature**

Features:

* L190300 - RBC
* FIELD\_29 - family history (hypertension)
* S000300 - BMI
* FIELD\_38 - Drinking 5-1 days (1 week)
* AGE
* SEX
* L100700 - Uric acid
* L100800 - Fasting glucose
* FIELD\_33 - 4-1 Smoking
* FIELD\_41 - 6-2 of physical activity (moderate)
* FIELD\_42 - 6-3 of physical activity (walking)
* L190500 - HCT
* L101600 - alkaline phosphatase
* L103100 - HDL-Cholesterol
* L103000 - Triglyceride
* DBP

R2 score 0.652

Random Forest Regressor

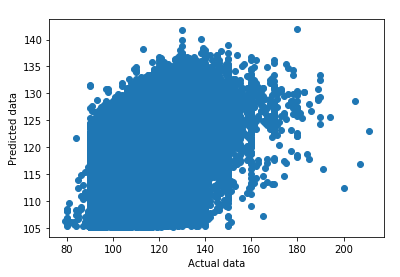


* 1. **Regression experiment excluding DBP as a feature**

Features:

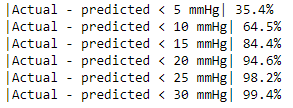
* 1. L190300 - RBC
  2. FIELD\_29 - family history (hypertension)
  3. S000300 - BMI
  4. FIELD\_38 - Drinking 5-1 days (1 week)
  5. AGE
  6. SEX
  7. L100700 - Uric acid
  8. L100800 - Fasting glucose
  9. FIELD\_33 - 4-1 Smoking
  10. FIELD\_41 - 6-2 of physical activity (moderate)
  11. FIELD\_42 - 6-3 of physical activity (walking)
  12. L190500 - HCT
  13. L101600 - alkaline phosphatase
  14. L103100 - HDL-Cholesterol
  15. L103000 - Triglyceride

Results:



R2 score 0.298

Random Forest Regressor

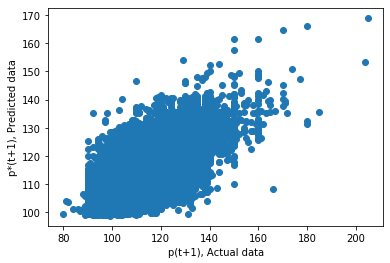


* 1. **Regression experiment for predicting next year SBP values**

Features:

1. S000501 - SBP
2. S000300 - BMI
3. S000502 - DBP
4. L103000 - Triglyceride
5. L100800 - Fasting glucose
6. L103100 - HDL-Cholesterol
7. L100700 - Uric Acid
8. L101300 - SGPT
9. L190500 - HCT
10. L190300 - RBC
11. L103300 - Cardiac risk factor
12. L190400 - Hemoglobin
13. L101700 - r-GTP gamma
14. L101600 - alkaline phosphatase

Results:



R2 score 0.488

Random Forest Regressor

