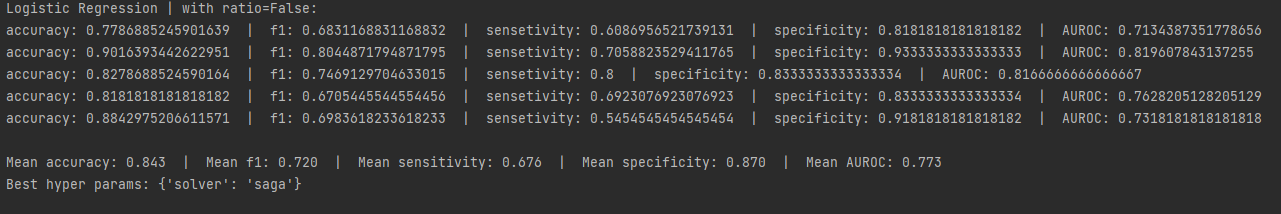
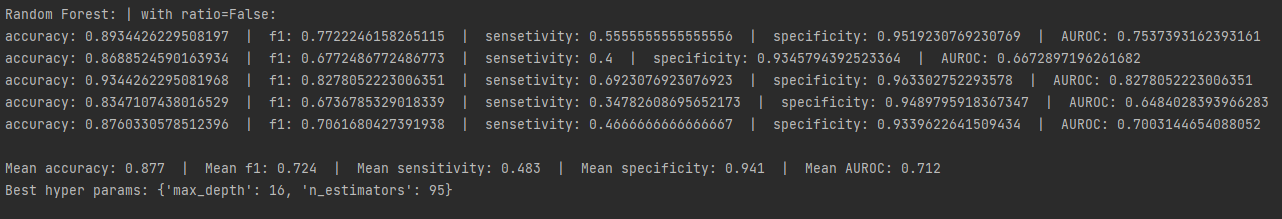
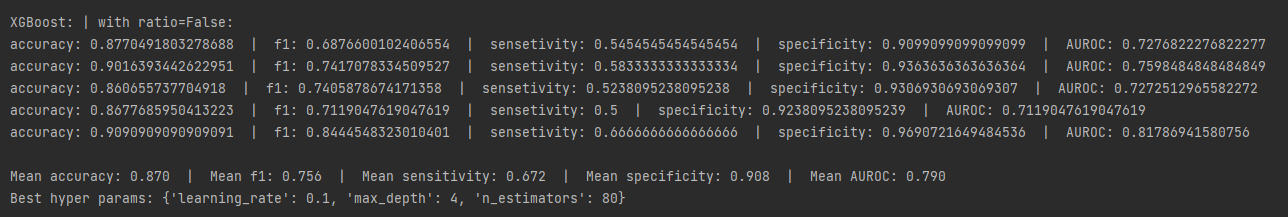
**Task 1 – Predict COVID-19 from routine blood tests:**

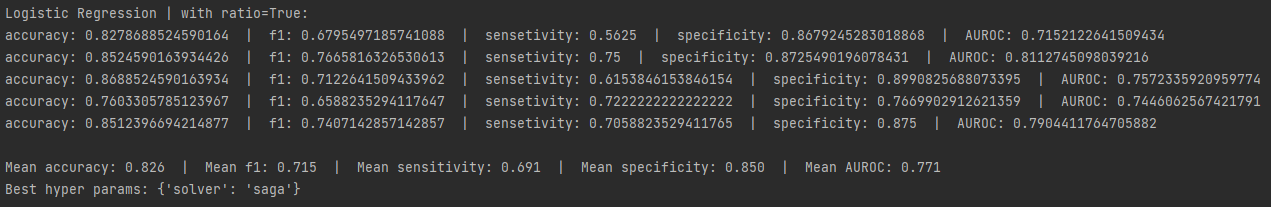
3. - The first new feature will be the ratio between Hemoglobin and Red blood cells () because they are strongly related with the amount of oxygen in the blood.  
 - The second new feature will be the ratio between Leukocytes and Red blood cells () which is the ratio between white blood cells (the type that fight against infections) and red blood cells.  
 - Another new feature will be the ratio between Lymphocytes and Red blood cells (), which is the ratio between white and red blood cells.  
 - Another new feature will be the ration between Monocytes and Red blood cells () which is the ratio between immune cells and red blood cells.  
 - Another new feature will be the ration between Neutrophils and Red blood cells () which is the ratio between immune cells and red blood cells.  
**Results:**

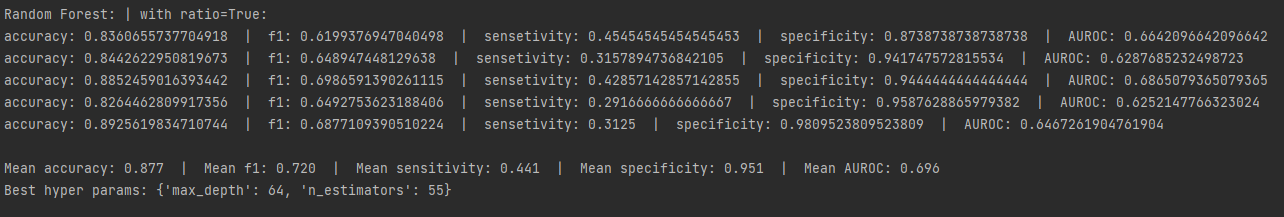
Results without new features:  


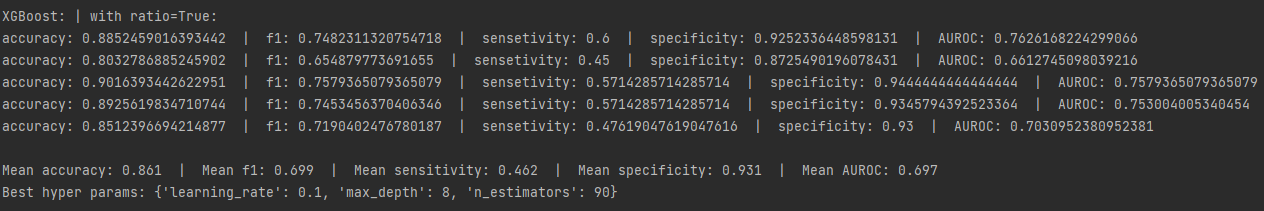




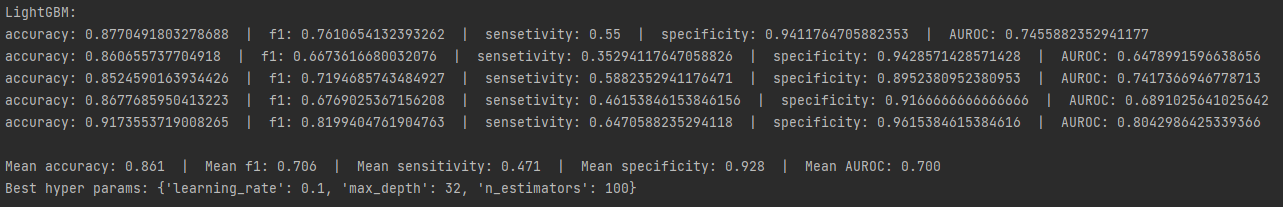
Results with new features:







The mean accuracy rate of the models are +-2% deviation from the results without the new features. However, the hyper-parameters changed.

4. Results(Evaluations for LightGMB and CatBoost, evaluations for Logistic Regression, Random Forest and XGBoost can be found in the previous question at ‘Results without new features’):  
LightGMB:  


CatBoost:

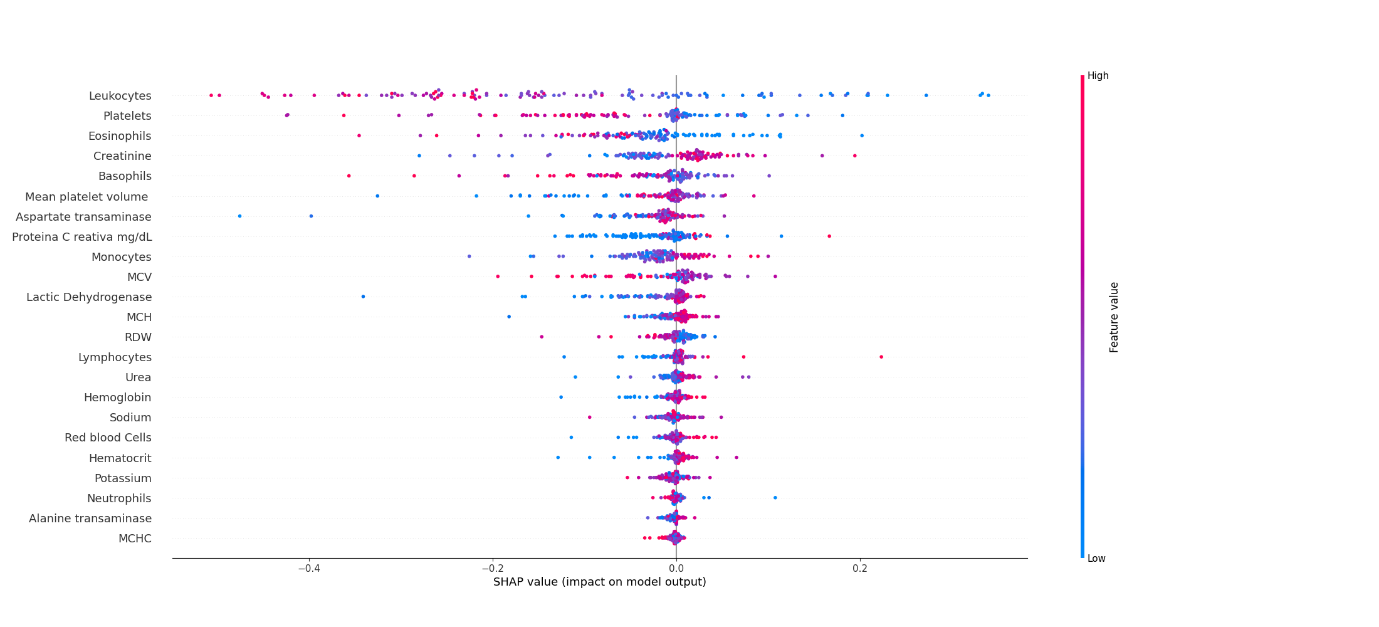


Assumption: The CatBoost evaluation was made with a change to the max\_depth hyper parameter

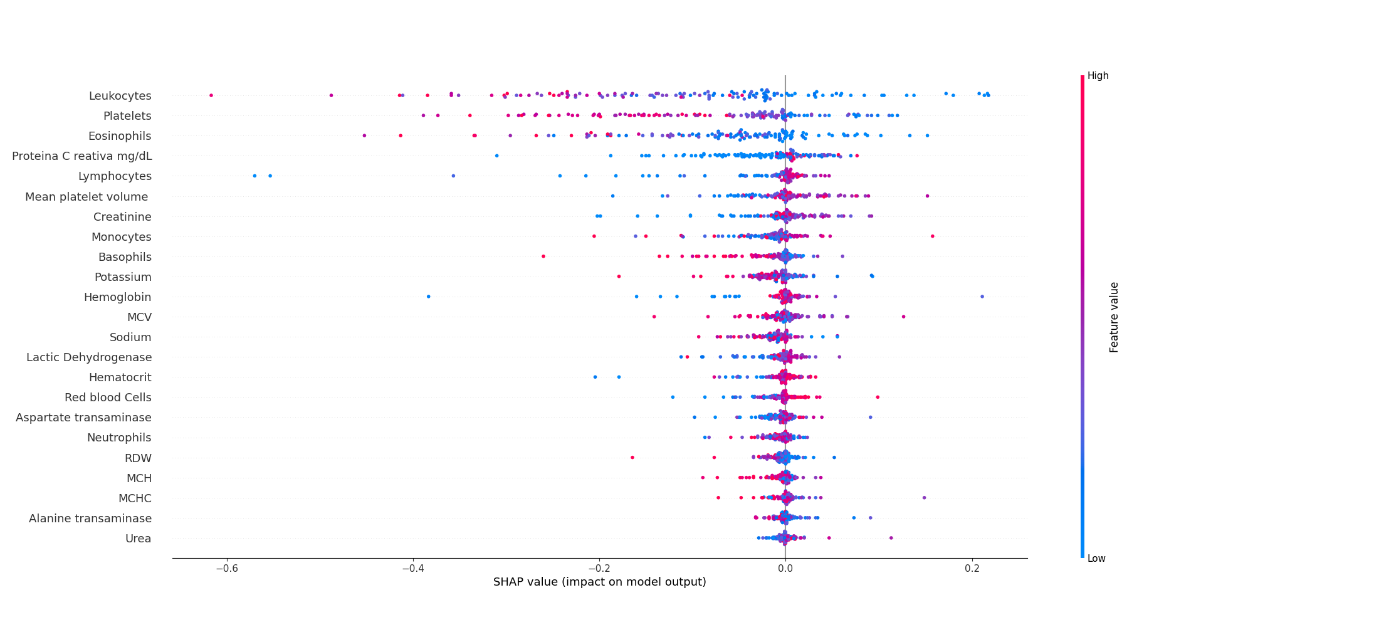
With values [2, 4, 8] and not [2, 4, 8, 16, 32, 64] since it resulted in unreasonable running time of multiple hours.

6. SHAP Illustrations:

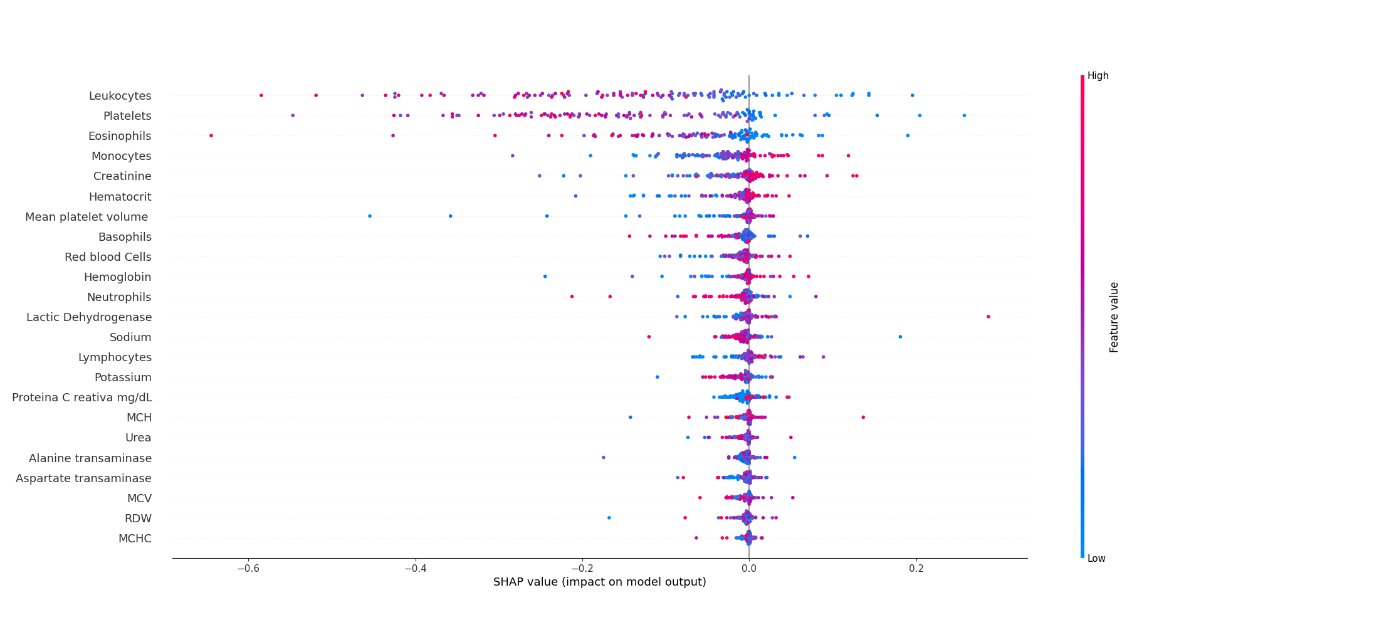
XGBoost:



LightGBM:



Random forest:



CatBoost:

