BS6200 Project

Epileptic-Seizure Classification

NI YUXIN

Content

Introduction

Preprocessing

Metrics Selection

Model evaluation

Summary and Analysis

Introduction

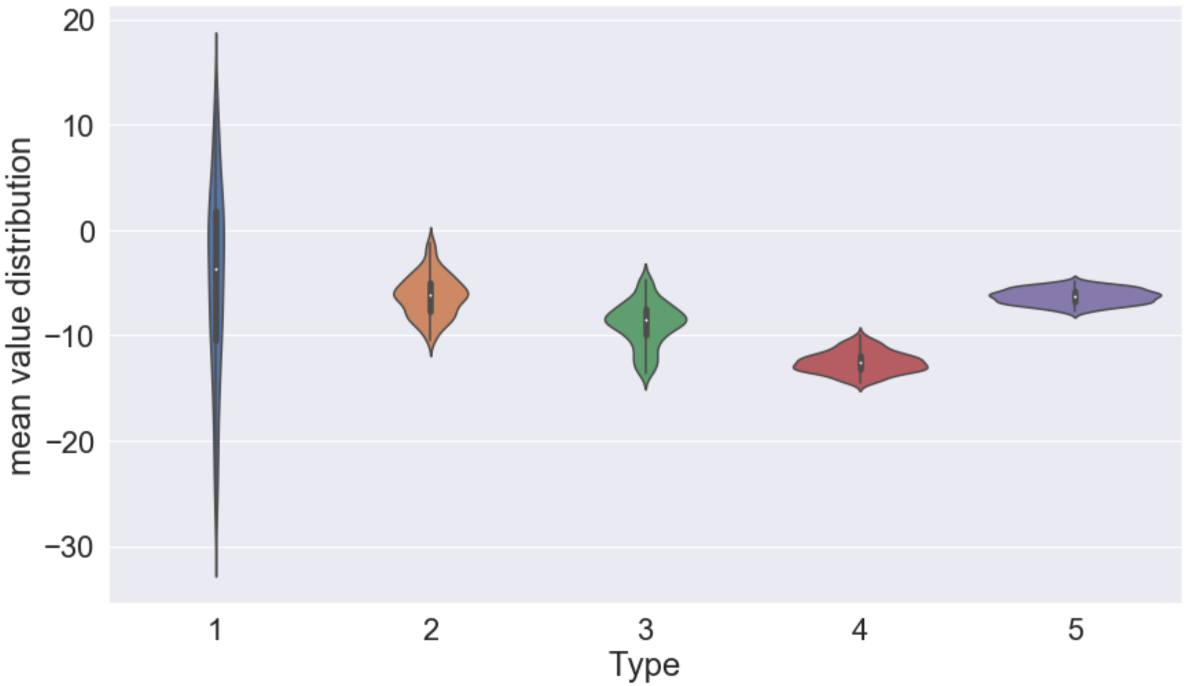
A seizure is an abnormal electrical discharge of a group of brain cells. It can cause different symptoms, depending on the location of the seizure and the spread of electrical activity through the brain.

Epilepsy impacts approximately 50 million people worldwide with an estimated annual cost of $12.5 billion for patients in the United States.

Research[1] suggest that seizures can have a direct adverse effect on cognition

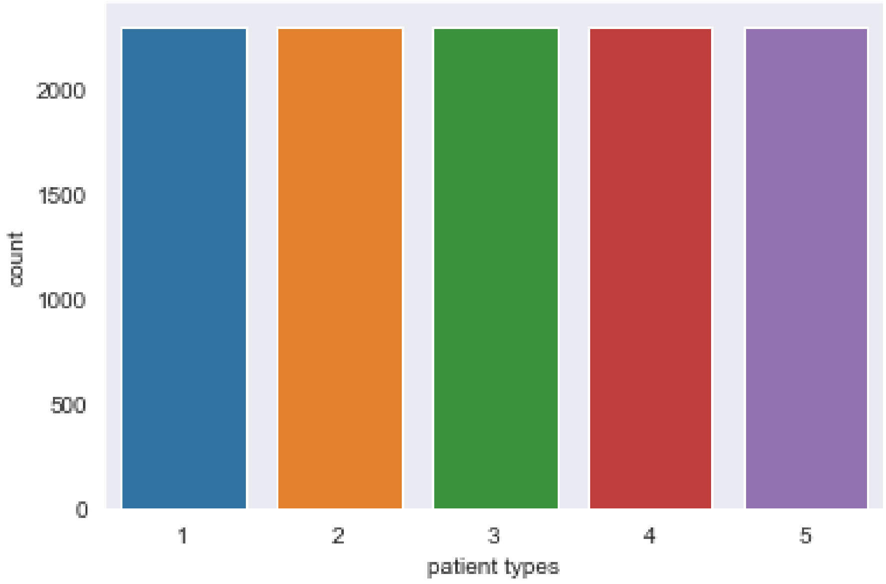
[1] Bergen, Donna C. "Do seizures harm the brain?." *Epilepsy currents* 6.4 (2006): 117-118.

EEG distribution

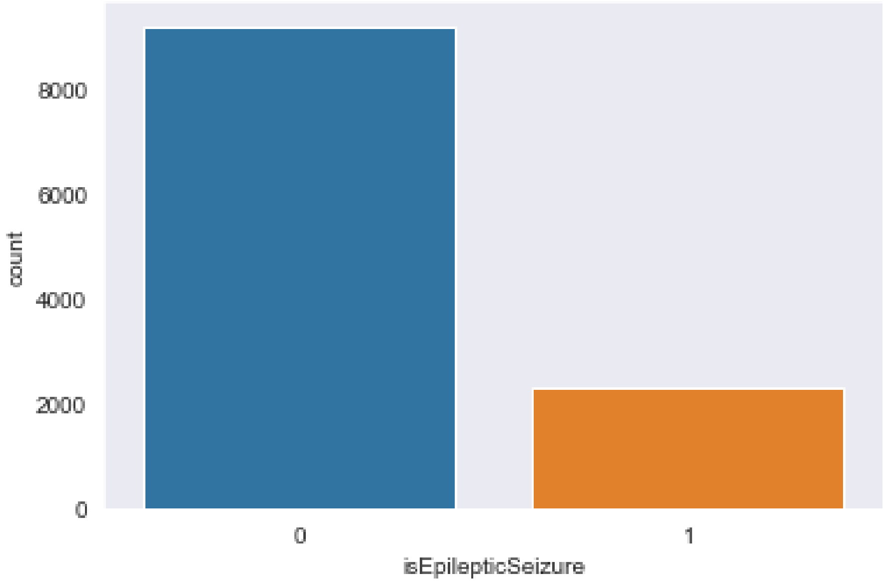


EEG in patient is very different.

Data distribution



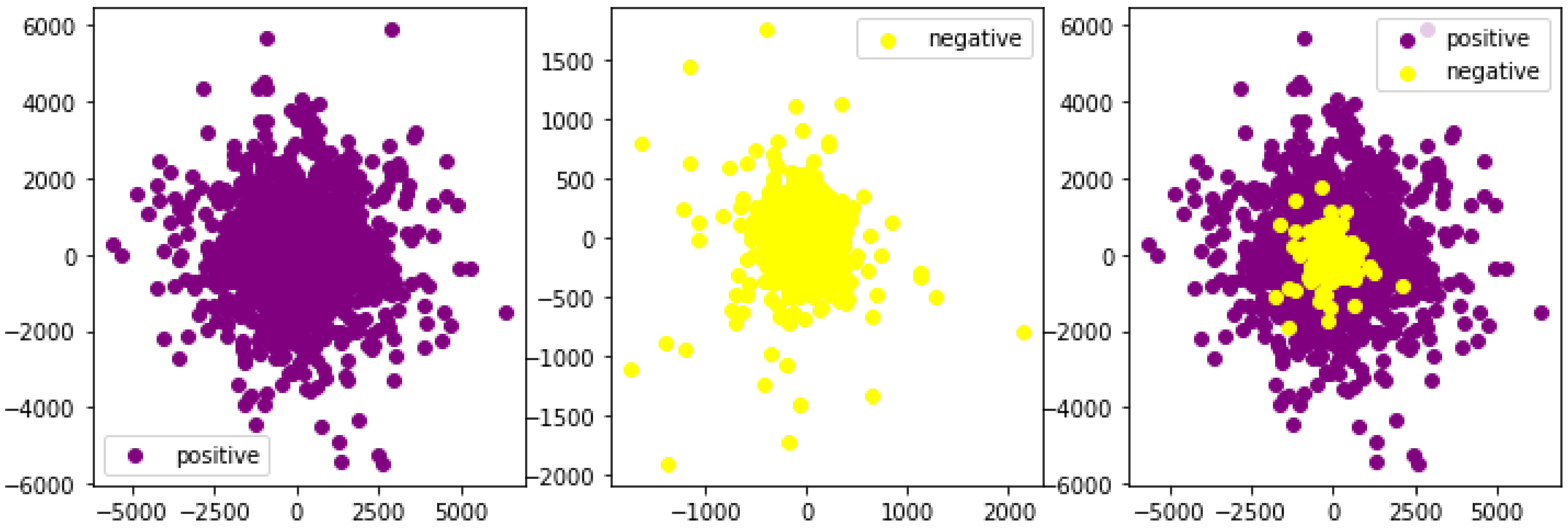
Five labels Data



Has Seizure Data

Five labels data is balanced while binary data is imbalanced

PCA Visualization



Distribution of negative case are more convergent.

Content

Introduction

Preprocessing

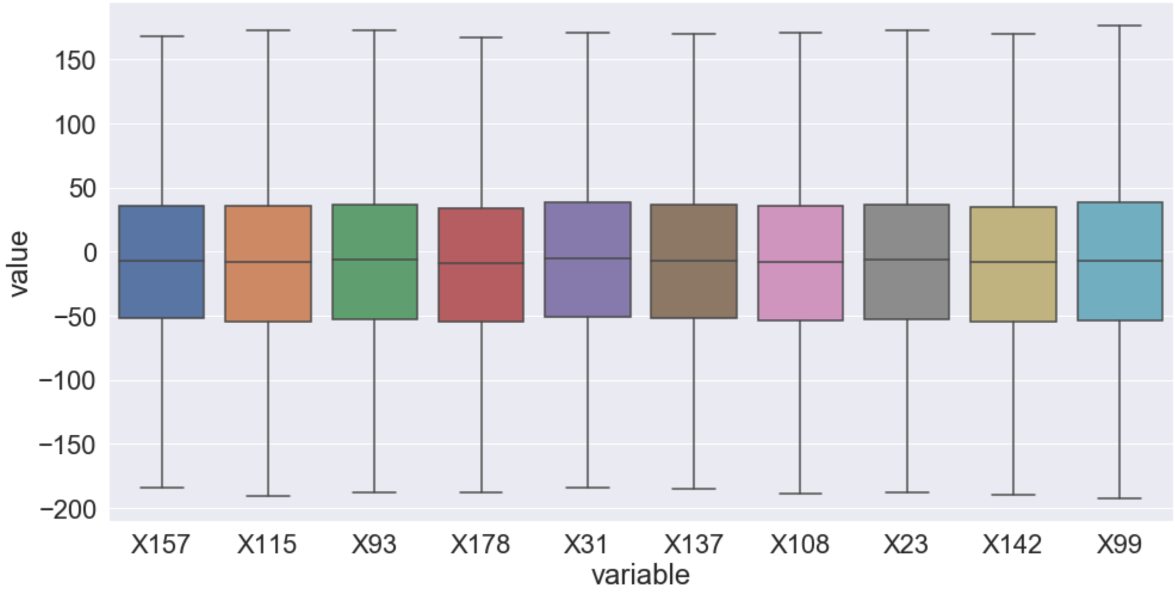
Metrics Selection

Model evaluation

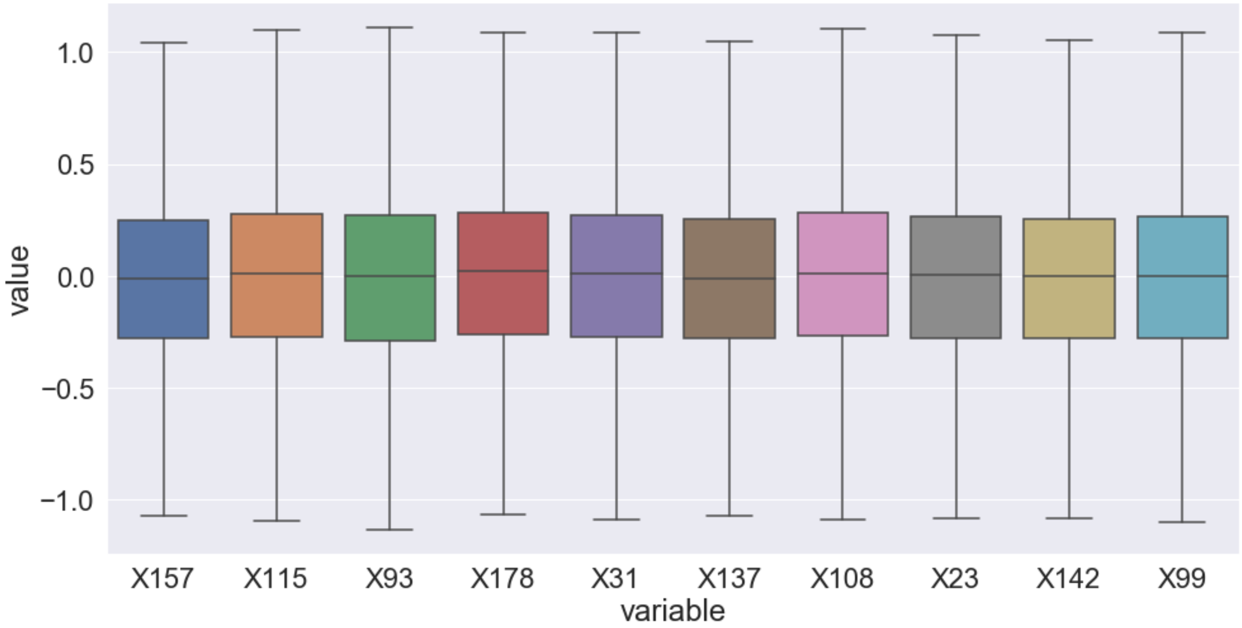
Summary and Analysis

Normalization

Normalization method $x_n = \dfrac{x - \mu}{\sigma}$



Before Normalization



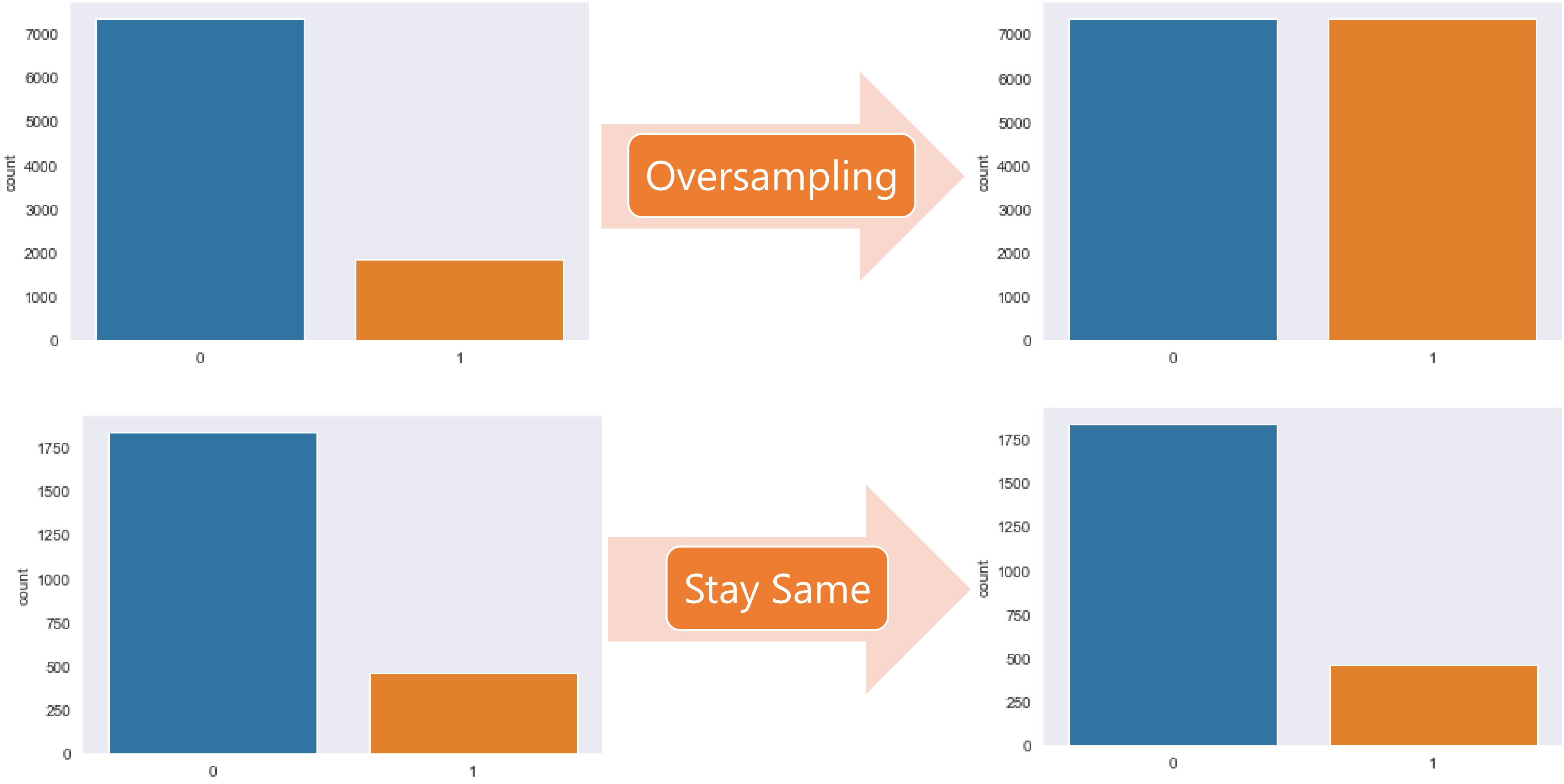
After Normalization

Split data and oversampling

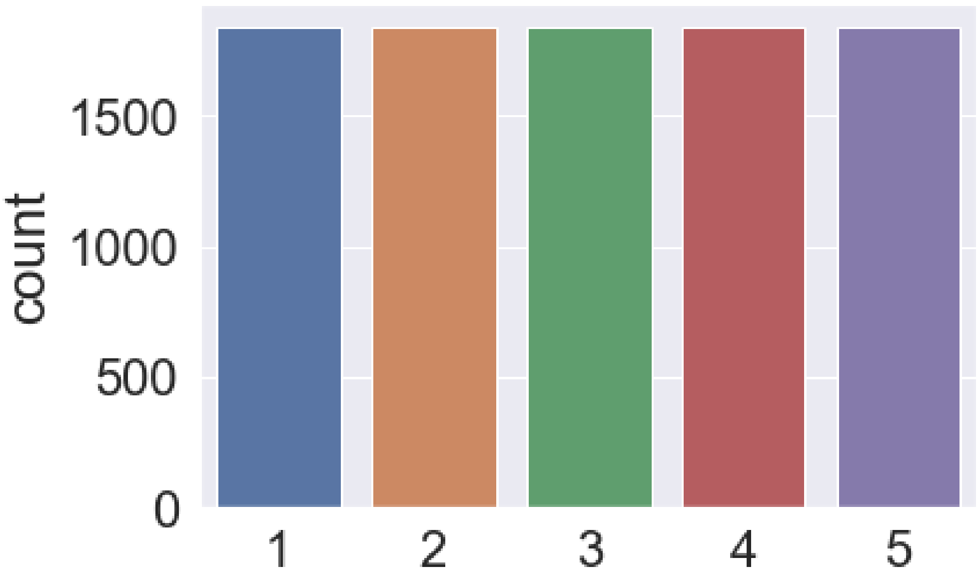
We split the total data to 80% train and 20% test data, and then use ADASYN to oversample our train data

Train Data

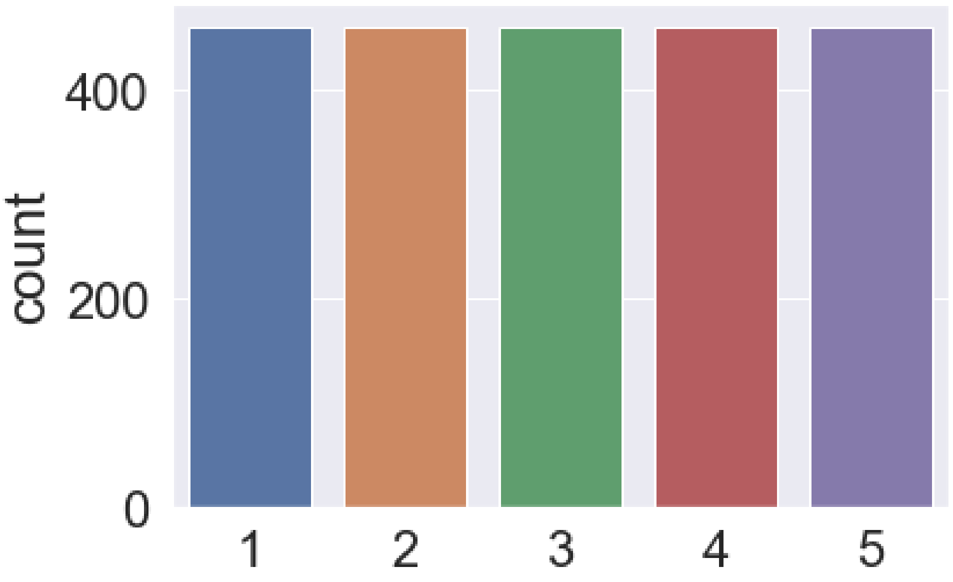
Test Data



Split data and oversampling



Train Data



Test Data

Content

Introduction

Preprocessing

Metrics Selection

Model evaluation

Summary and Analysis

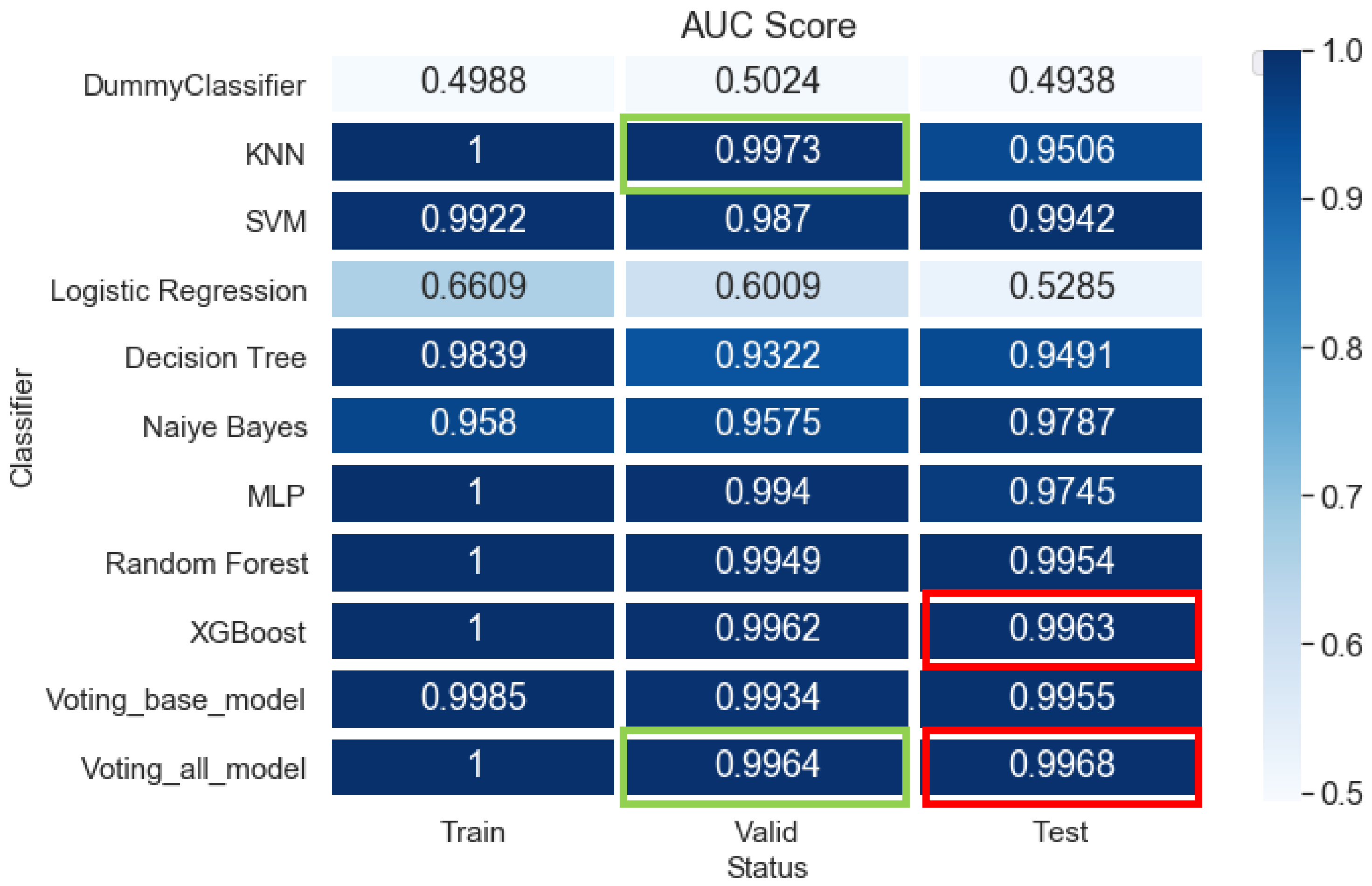
Goal Statement in binary classification

Predict Epileptic Seizure in imbalanced data

Focus more on performance towards patient

F1 Score in patient

ROC AUC Performance

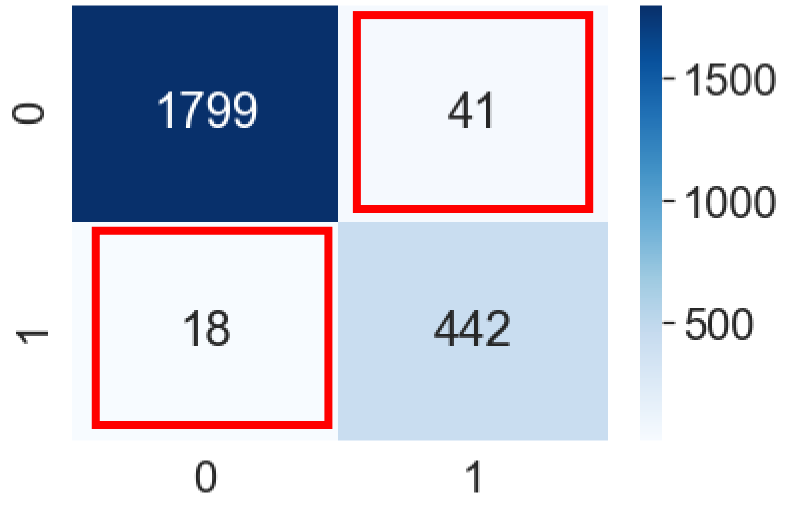


KNN, Decision Tree ,Random Forest and XGBoost are fined tuned by 5-fold cross validation

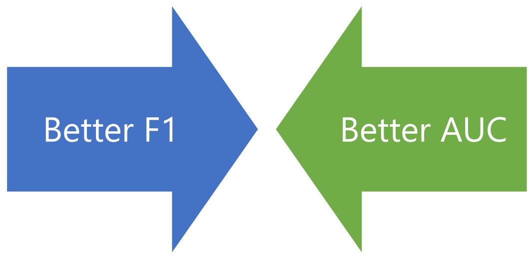
Voting by all models is the best model and XGBoost is the second

F1 VS AUC

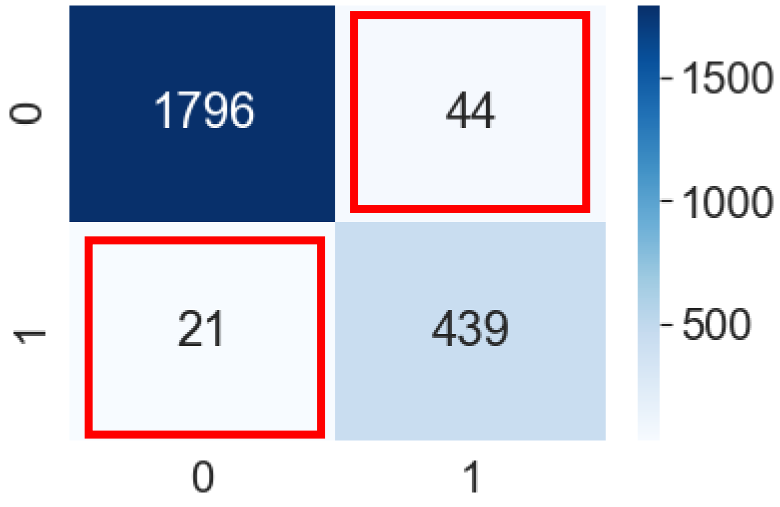
XGBoost



AUC: 0.9963



Voting Classifier with all models



AUC: 0.9968

ROC AUC is the area under ROC curve, it is the average value of all the thresholds

F1 score is the value under a specific threshold, here is 0.5, the threshold may not be right

ROC AUC is optimistic in imbalanced data

ROC AUC is the area under ROC curve, it is the average value of all the thresholds and treats equally to majority and minority

> *ROC analysis does not have any bias toward models that perform well on the minority class at the expense of the majority class—a property that is quite attractive when dealing with imbalanced data.*

— Page 27, Imbalanced Learning: Foundations, Algorithms, and Applications, 2013.

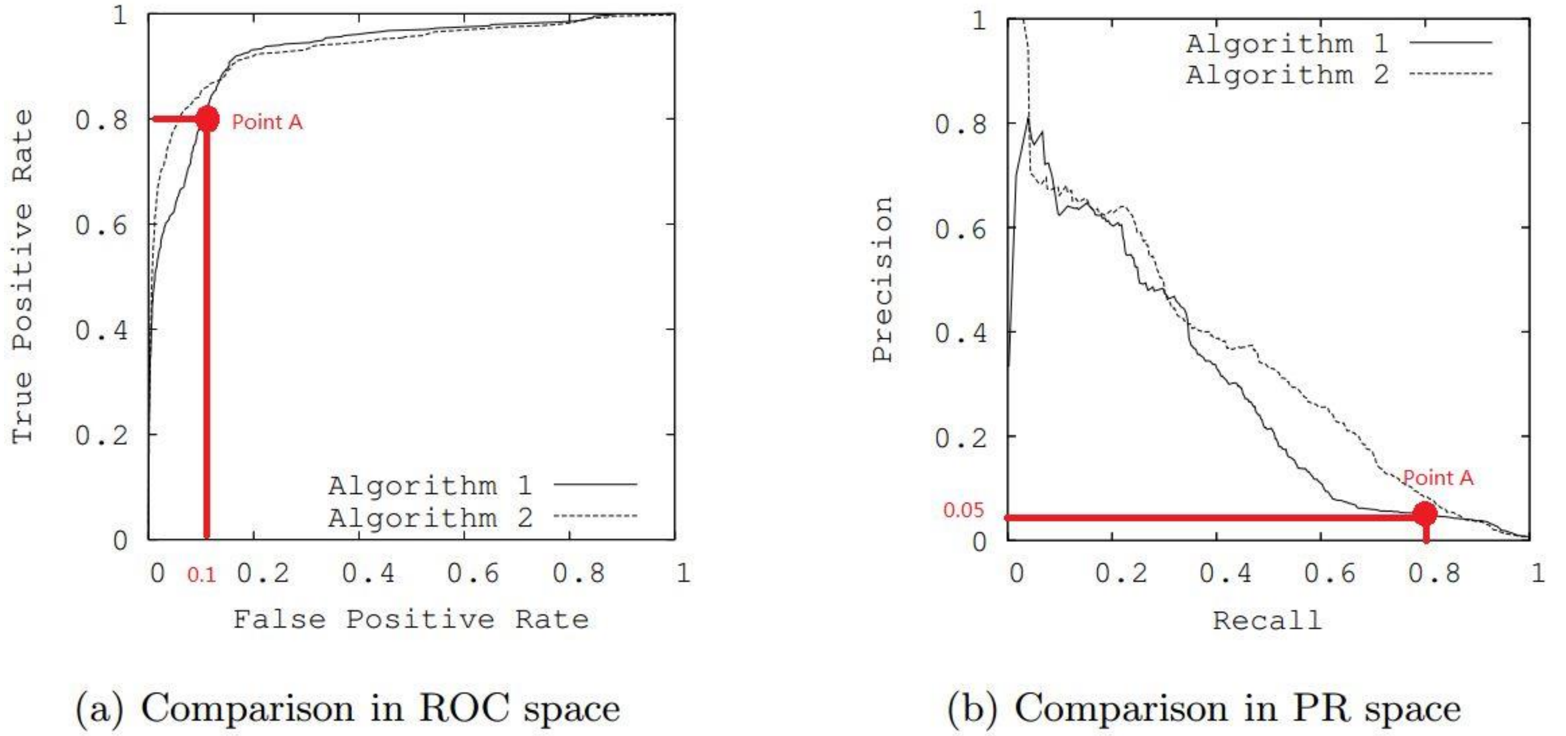
> *Although ROC graphs are widely used to evaluate classifiers under presence of class imbalance, it has a drawback: under class rarity, that is, when the problem of class imbalance is associated to the presence of a low sample size of minority instances, as the estimates can be unreliable.*

— Page 55, Learning from Imbalanced Data Sets, 2018.

> *Precision-recall curves (PR curves) are recommended for highly skewed domains where ROC curves may provide an excessively optimistic view of the performance.*

— A Survey of Predictive Modelling under Imbalanced Distributions, 2015.

PR AUC is more sensitive



For imbalanced data when positive data is less than negative data, **PR AUC is more suitable.**

Davis, Jesse, and Mark Goadrich. "The relationship between Precision-Recall and ROC curves." *Proceedings of the 23rd international conference on Machine learning*. 2006. **Cited by 4384**

Metrics strategy

Binary:

Find model based on PR AUC

Find The Best Threshold based on PR Curve

Best model

Comparison Strategy

First Compare PR AUC

Then compare F1 score under best threshold

Finally compare recall

Multi-class: ROC AUC and F1 Score are chosen as the metrics in Multi-class classification

Content

Introduction

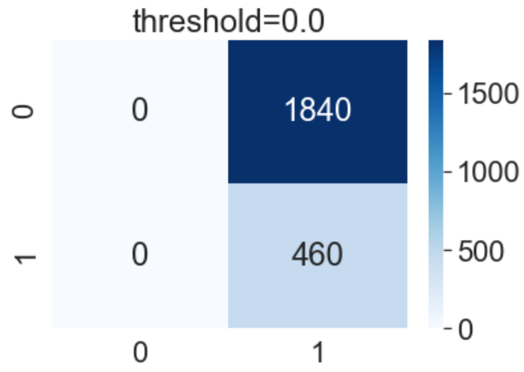
Preprocessing

Metrics Selection

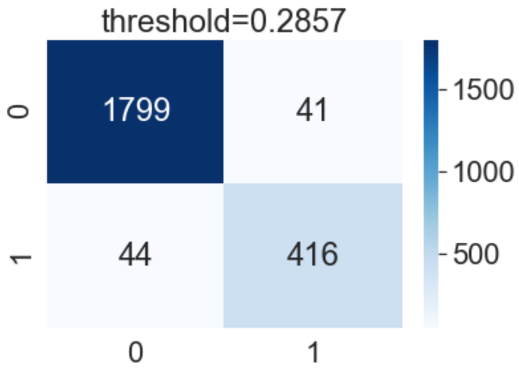
Model evaluation

Summary and Analysis

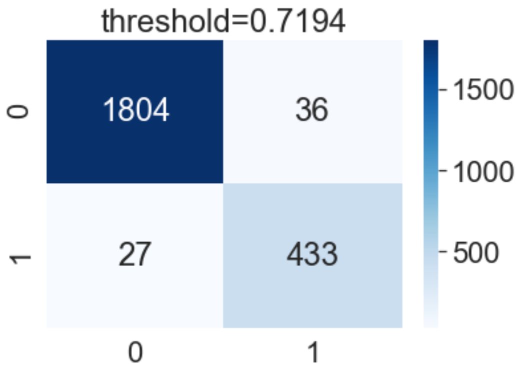
Base Model-Test Performance



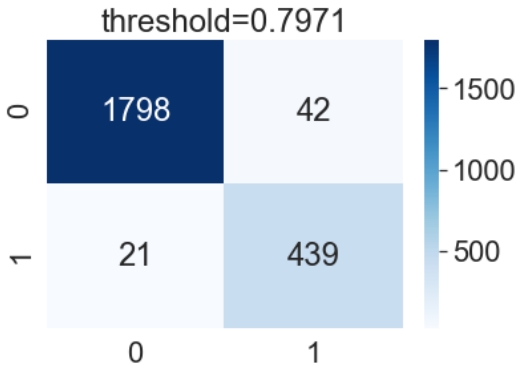
Dummy clf



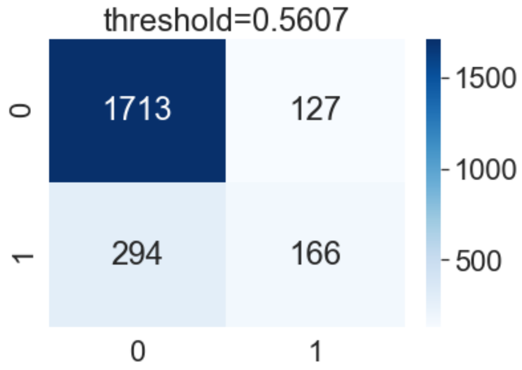
KNN



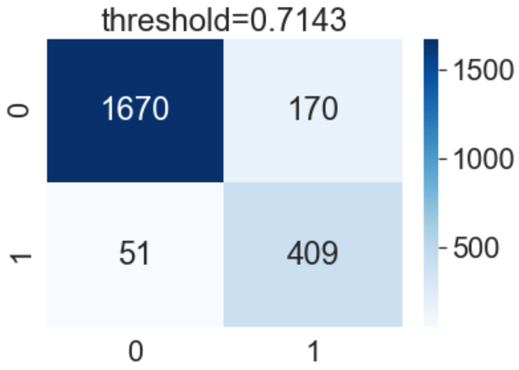
SVM



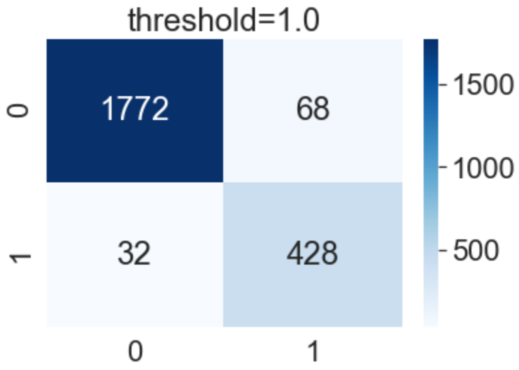
4-layers DNN



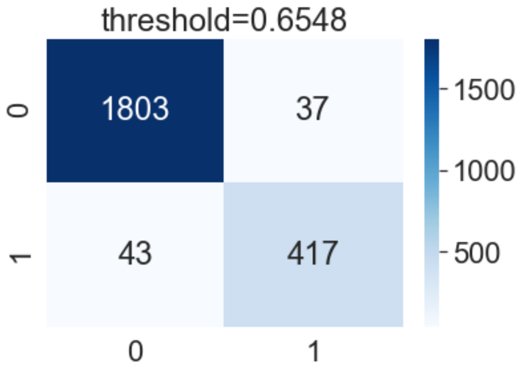
LR



Decision Tree



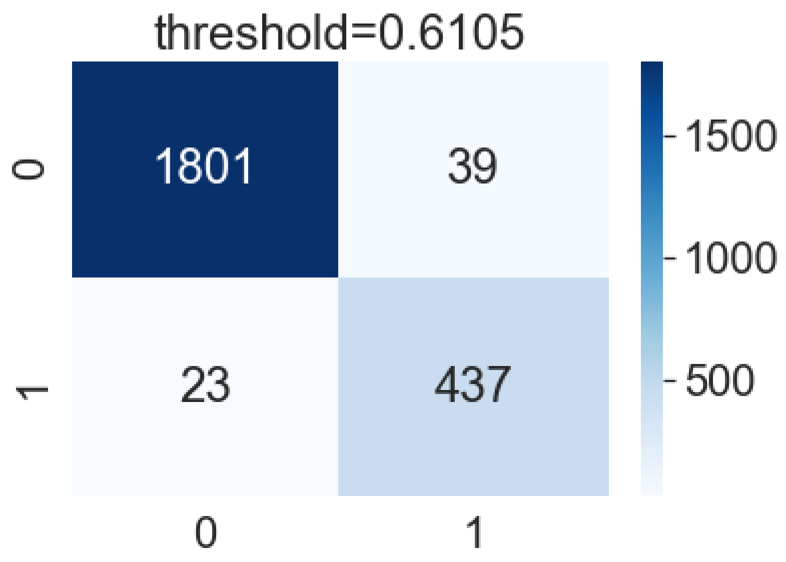
Naïve Bayes



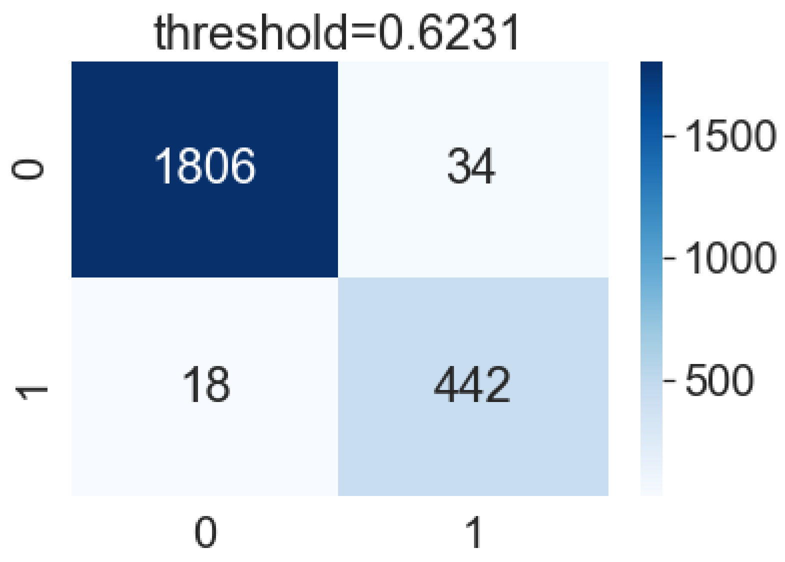
MLP

Ensemble Learning- Test Performance

Random Forest



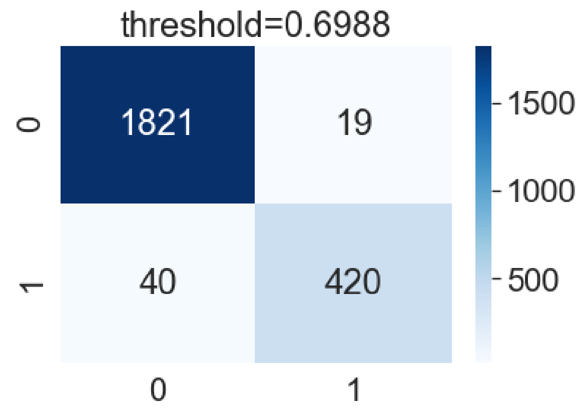
XGBoost



Voting Classifier

Model
KNN
SVM
Decision Tree
Naïve Bayes
MLP

[4, 8, 1, 2, 4]



PRAUC:0.9845

Better than all base models and Random Forest

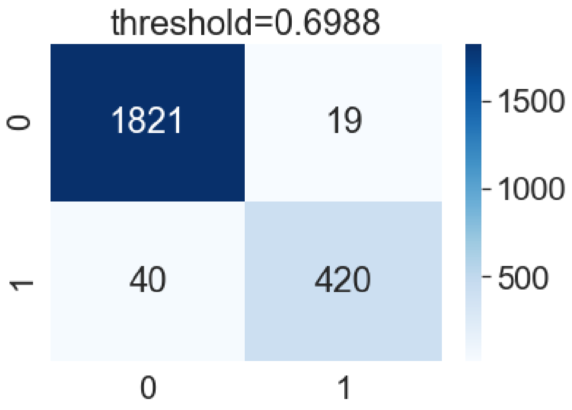
$$\boxed{Weight = \frac{Test\_data\_PR\_AUC}{1 - Test\_data\_PR\_AUC}}$$

Big values get more bigger

Voting Classifier

Model
KNN
SVM
Decision Tree
Naïve Bayes
MLP

[4, 8, 1, 2, 4]



PRAUC:0.9845

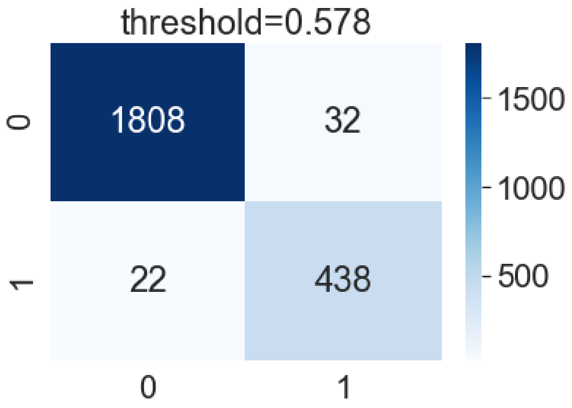
Better than all base models and Random Forest

Model
KNN
SVM
Decision Tree
Naïve Bayes
MLP
Random Forest
XGBoost

$$Weight = \frac{Test\_data\_PR\_AUC}{1 - Test\_data\_PR\_AUC}$$

Big values get more bigger

[4, 8, 1, 2, 4, 9, 11]



Better than all other models

PRAUC:0.98878

Result

Model	Fine tune?	hyperparameter	Best threshold for patient	Average Weighted F1 Score		Patient F1 (best)	ROCAUC	PRAUC
				threshold=0.5	threshold=best threshold			
KNN	Y	k=7	0.2857	0.9628	0.963	0.907	0.958	0.957
SVM	Y	kernel="rbf"	0.7194	0.970	0.9722	0.932	0.994	0.979
LR	N		0.5607	0.6828	0.8	0.441	0.528	0.438
Decision Tree	Y	max depth=50,min sample split=100	0.7143	0.8919	0.9069	0.787	0.9485	0.849
Naïve Bayes	N		1	0.9536	0.9566	0.8954	0.9786	0.92271
MLP	Y	hidden layers=(256,256)	0.6548	0.9644	0.9651	0.9124	0.9774	0.95982
Random Forest	Y	n_estimators=100, min_samples_split =4	0.5717	0.9619	0.9712	0.930	0.9953	0.9792
XGBoost	Y	n estimators=150, max_depth=4,gamma=0	0.6231	0.9759	0.9771	0.9444	0.9962	0.9856
Voting-with-base-model-1	N		0.7865	0.9699	0.9728	0.9343	0.9957	0.9845
Voting-with-all-models-2	N		0.578	0.97416	0.9766	0.9419	0.9970	0.98878

Result

Model	Fine tune?	hyperparameter	Best threshold for patient	Average Weighted F1 Score		Patient F1 (best)	ROCAUC	PRAUC
				threshold=0.5	threshold=best threshold			
KNN	Y	k=7	0.2857	0.9628	0.963	0.907	0.958	0.957
SVM	Y	kernel="rbf"	0.7194	0.970	0.9722	0.932	0.994	0.979
LR	N		0.5607	0.6828	0.8	0.441	0.528	0.438
Decision Tree	Y	max depth=50,min sample split=100	0.7143	0.8919	0.9069	0.787	0.9485	0.849
Naïve Bayes	N		1	0.9536	0.9566	0.8954	0.9786	0.92271
MLP	Y	hidden layers=(256,256)	0.6548	0.9644	0.9651	0.9124	0.9774	0.95982
Random Forest	Y	n_estimators=100, min_samples_split=4	0.5717	0.9619	0.9712	0.930	0.9953	0.9792
XGBoost	Y	n estimators=150, max_depth=4,gamma=0	0.6231	0.9759	0.9771	0.9444	0.9962	0.9856
Voting-with-base-model-1	N		0.7865	0.9699	0.9728	0.9343	0.9957	0.9845
Voting-with-all-models-2	N		0.578	0.97416	0.9766	0.9419	0.9970	0.98878

Result

Model	Fine tune?	hyperparameter	Best threshold for patient	Average Weighted F1 Score		Patient F1 (best)	ROCAUC	PRAUC
				threshold=0.5	threshold=best threshold			
KNN	Y	k=7	0.2857	0.9628	0.963	0.907	0.958	0.957
SVM	Y	kernel="rbf”	0.7194	0.970	0.9722	0.932	0.994	0.979
LR	N		0.5607	0.6828	0.8	0.441	0.528	0.438
Decision Tree	Y	max depth=50,min sample split=100	0.7143	0.8919	0.9069	0.787	0.9485	0.849
Naïve Bayes	N		1	0.9536	0.9566	0.8954	0.9786	0.92271
MLP	Y	hidden layers=(256,256)	0.6548	0.9644	0.9651	0.9124	0.9774	0.95982
Random Forest	Y	n_estimators=100, min_samples_split =4	0.5717	0.9619	0.9712	0.930	0.9953	0.9792
XGBoost	Y	n estimators=150, max_depth=4,gamma=0	0.6231	0.9759	0.9771	0.9444	0.9962	0.9856
Voting-with-base-model-1	N		0.7865	0.9699	0.9728	0.9343	0.9957	0.9845
Voting-with-all-models-2	N		0.578	0.97416	0.9766	0.9419	0.9970	0.98878

Result

Model	Fine tune?	hyperparameter	Best threshold for patient	Average Weighted F1 Score		Patient F1 (best)	ROCAUC	PRAUC
				threshold=0.5	threshold=best threshold			
KNN	Y	k=7	0.2857	0.9628	0.963	0.907	0.958	0.957
SVM	Y	kernel="rbf"	0.7194	0.970	0.9722	0.932	0.994	0.979
LR	N		0.5607	0.6828	0.8	0.441	0.528	0.438
Decision Tree	Y	max depth=50,min sample split=100	0.7143	0.8919	0.9069	0.787	0.9485	0.849
Naïve Bayes	N		1	0.9536	0.9566	0.8954	0.9786	0.92271
MLP	Y	hidden layers=(256,256)	0.6548	0.9644	0.9651	0.9124	0.9774	0.95982
Random Forest	Y	n_estimators=100, min_samples_split =4	0.5717	0.9619	0.9712	0.930	0.9953	0.9792
XGBoost	Y	n estimators=150, max_depth=4,gamma=0	0.6231	0.9759	0.9771	0.9444	0.9962	0.9856
Voting-with-base-model-1	N		0.7865	0.9699	0.9728	0.9343	0.9957	0.9845
Voting-with-all-models-2	N		0.578	0.97416	0.9766	0.9419	0.9970	0.98878

Result

Model	Fine tune?	hyperparameter	Best threshold for patient	Average Weighted F1 Score		Patient F1 (best)	ROCAUC	PRAUC
				threshold=0.5	threshold=best threshold			
KNN	Y	k=7	0.2857	0.9628	0.963	0.907	0.958	0.957
SVM	Y	kernel="rbf"	0.7194	0.970	0.9722	0.932	0.994	0.979
LR	N		0.5607	0.6828	0.8	0.441	0.528	0.438
Decision Tree	Y	max depth=50,min sample split=100	0.7143	0.8919	0.9069	0.787	0.9485	0.849
Naïve Bayes	N		1	0.9536	0.9566	0.8954	0.9786	0.92271
MLP	Y	hidden layers=(256,256)	0.6548	0.9644	0.9651	0.9124	0.9774	0.95982
Random Forest	Y	n_estimators=100, min_samples_split =4	0.5717	0.9619	0.9712	0.930	0.9953	0.9792
XGBoost	Y	n estimators=150, max_depth=4,gamma=0	0.6231	0.9759	0.9771	0.9444	0.9962	0.9856
Voting-with-base-model-1	N		0.7865	0.9699	0.9728	0.9343	0.9957	0.9845
Voting-with-all-models-2	N		0.578	0.97416	0.9766	0.9419	0.9970	0.98878

Result

Model	Fine tune?	hyperparameter	Best threshold for patient	Average Weighted F1 Score		Patient F1 (best)	ROCAUC	PRAUC
				threshold=0.5	threshold=best threshold			
KNN	Y	k=7	0.2857	0.9628	0.963	0.907	0.958	0.957
SVM	Y	kernel="rbf"	0.7194	0.970	0.9722	0.932	0.994	0.979
LR	N		0.5607	0.6828	0.8	0.441	0.528	0.438
Decision Tree	Y	max depth=50,min sample split=100	0.7143	0.8919	0.9069	0.787	0.9485	0.849
Naïve Bayes	N		1	0.9536	0.9566	0.8954	0.9786	0.92271
MLP	Y	hidden layers=(256,256)	0.6548	0.9644	0.9651	0.9124	0.9774	0.95982
Random Forest	Y	n_estimators=100, min_samples_split =4	0.5717	0.9619	0.9712	0.930	0.9953	0.9792
XGBoost	Y	n estimators=150, max_depth=4,gamma=0	0.6231	0.9759	0.9771	0.9444	0.9962	0.9856
Voting-with-base-model-1	N		0.7865	0.9699	0.9728	0.9343	0.9957	0.9845
Voting-with-all-models-2	N		0.578	0.97416	0.9766	0.9419	0.9970	0.98878

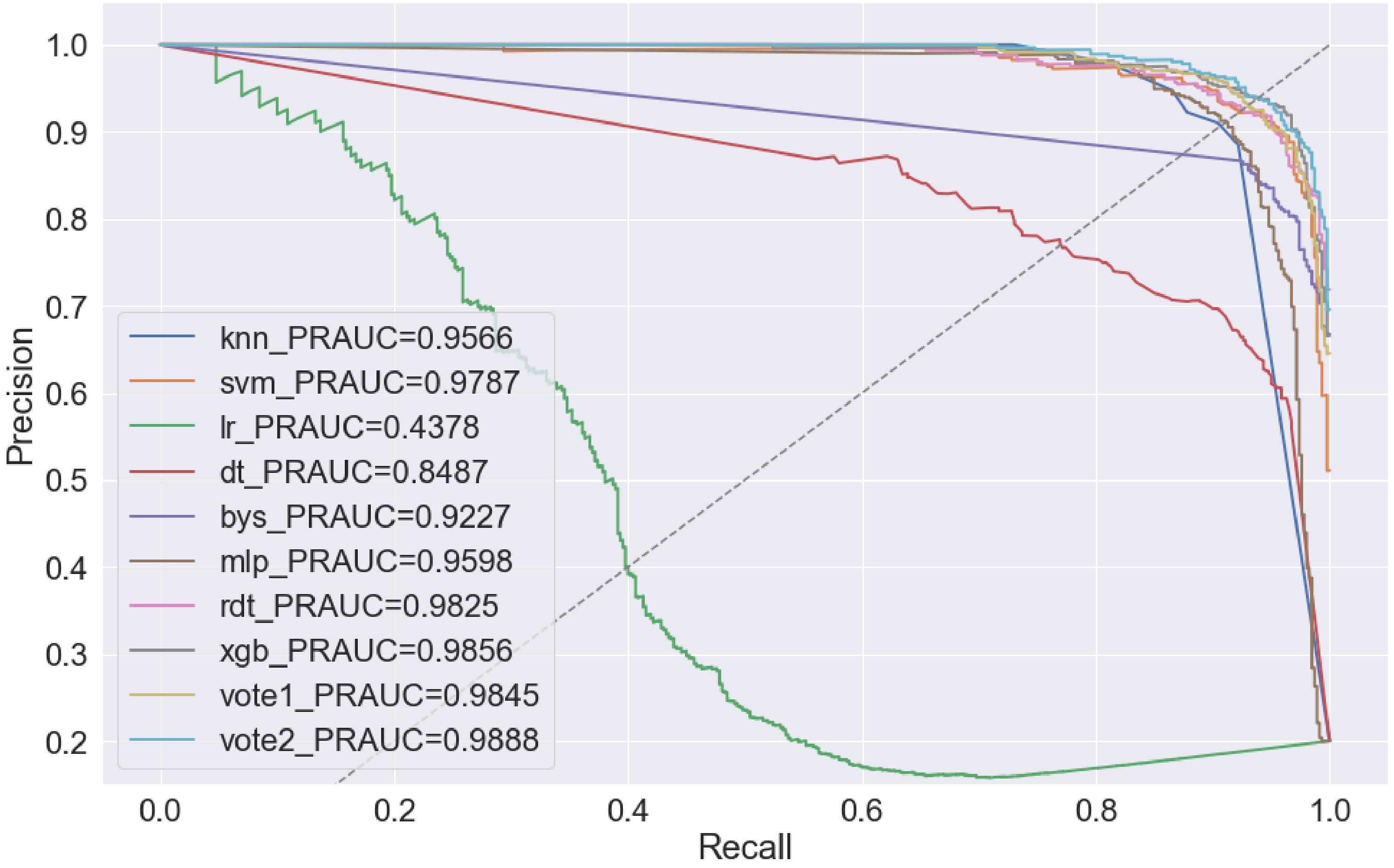
Result

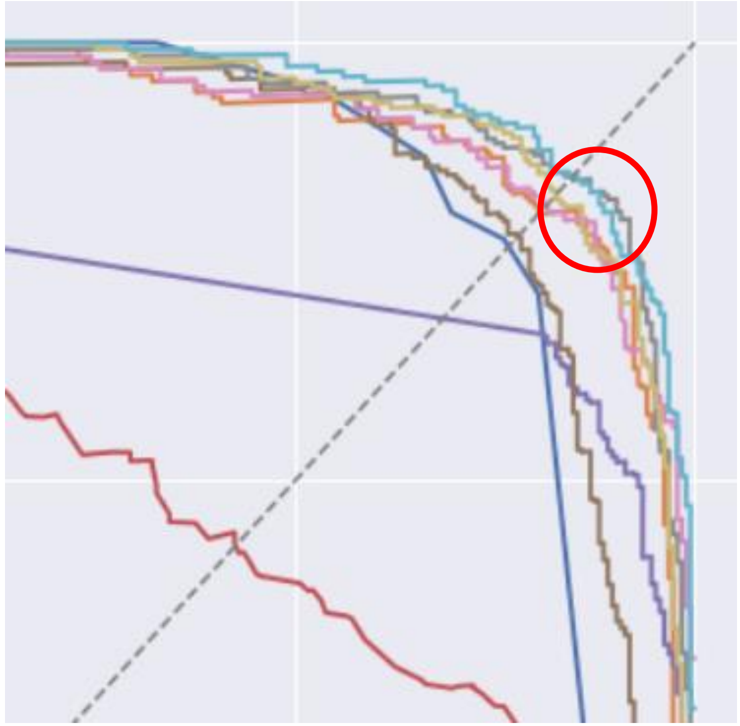
Model	Fine tune?	hyperparameter	Best threshold for patient	Average Weighted F1 Score		Patient F1 (best)	ROCAUC	PRAUC
				threshold=0.5	threshold=best threshold			
KNN	Y	k=7	0.2857	0.9628	0.963	0.907	0.958	0.957
SVM	Y	kernel="rbf"	0.7194	0.970	0.9722	0.932	0.994	0.979
LR	N		0.5607	0.6828	0.8	0.441	0.528	0.438
Decision Tree	Y	max depth=50,min sample split=100	0.7143	0.8919	0.9069	0.787	0.9485	0.849
Naïve Bayes	N		1	0.9536	0.9566	0.8954	0.9786	0.92271
MLP	Y	hidden layers=(256,256)	0.6548	0.9644	0.9651	0.9124	0.9774	0.95982
Random Forest	Y	n_estimators=100, min_samples_split =4	0.5717	0.9619	0.9712	0.930	0.9953	0.9792
XGBoost	Y	n estimators=150, max_depth=4,gamma=0	0.6231	0.9759	0.9771	0.9444	0.9962	0.9856
Voting-with-base-model-1	N		0.7865	0.9699	0.9728	0.9343	0.9957	0.9845
Voting-with-all-models-2	N		0.578	0.97416	0.9766	0.9419	0.9970	0.98878

Result

Model	Fine tune?	hyperparameter	Best threshold for patient	Average Weighted F1 Score		Patient F1 (best)	ROCAUC	PRAUC
				threshold=0.5	threshold=best threshold			
KNN	Y	k=7	0.2857	0.9628	0.963	0.907	0.958	0.957
SVM	Y	kernel="rbf"	0.7194	0.970	0.9722	0.932	0.994	0.979
LR	N		0.5607	0.6828	0.8	0.441	0.528	0.438
Decision Tree	Y	max depth=50,min sample split=100	0.7143	0.8919	0.9069	0.787	0.9485	0.849
Naïve Bayes	N		1	0.9536	0.9566	0.8954	0.9786	0.92271
MLP	Y	hidden layers=(256,256)	0.6548	0.9644	0.9651	0.9124	0.9774	0.95982
Random Forest	Y	n_estimators=100, min_samples_split =4	0.5717	0.9619	0.9712	0.930	0.9953	0.9792
XGBoost	Y	n estimators=150, max_depth=4,gamma=0	0.6231	0.9759	0.9771	0.9444	0.9962	0.9856
Voting-with-base-model-1	N		0.7865	0.9699	0.9728	0.9343	0.9957	0.9845
Voting-with-all-models-2	N		0.578	0.97416	0.9766	0.9419	0.9970	0.98878

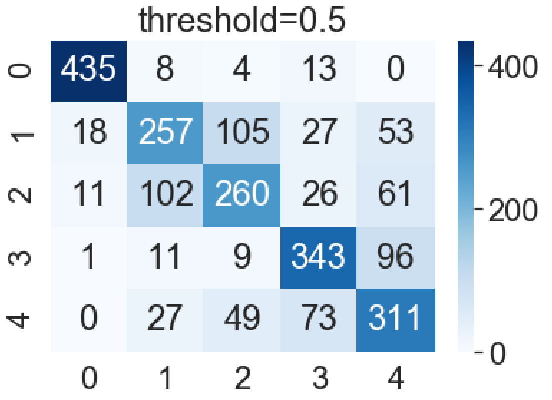
PR AUC Score



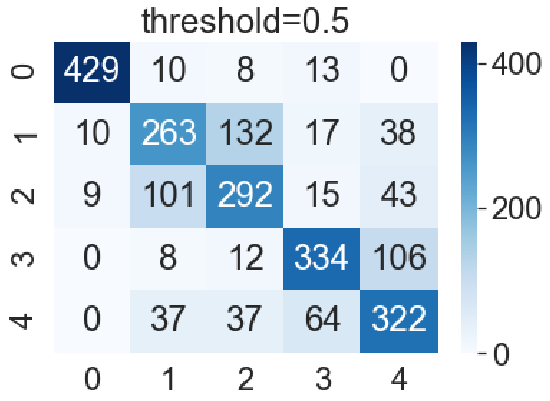


Voting classifier is best, globally

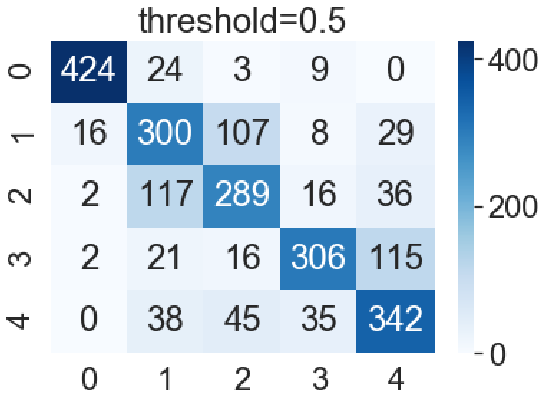
Performance of Multi-class



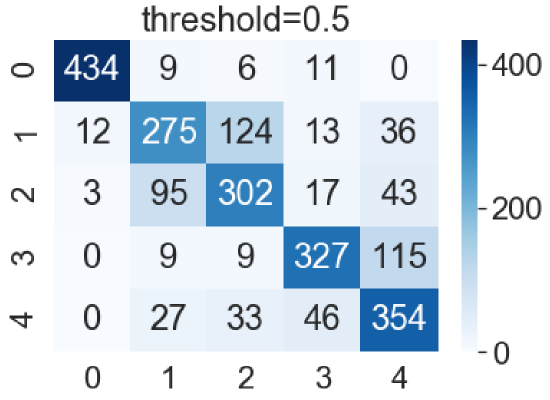
Random Forest



XGBoost



Voting with base models



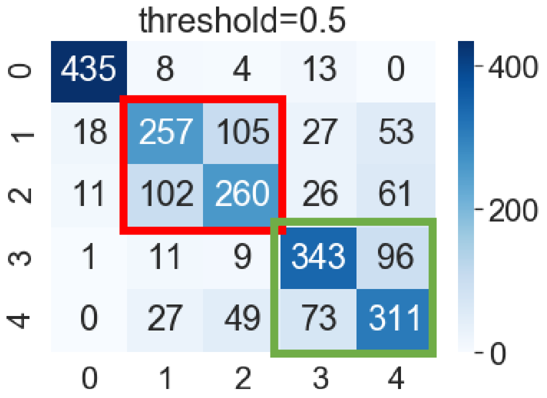
Voting with all models

Model	Average F1 Score	ROCAUC
Random Forest	0.697	0.916
XGBoost	0.714	0.924
Voting-with-base-models	0.724	0.909
Voting-with-all-models	0.736	0.9338

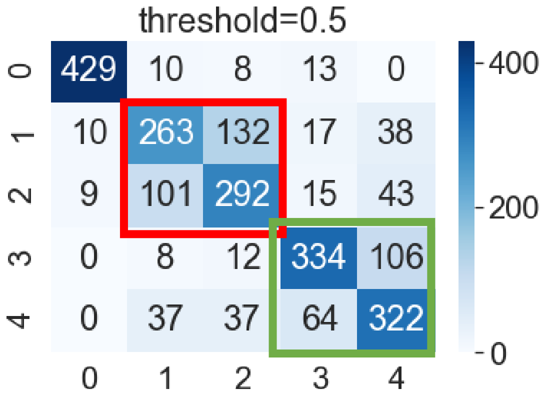
Voting classifier has largest F1 score and ROC AUC,

Xgboost is the second

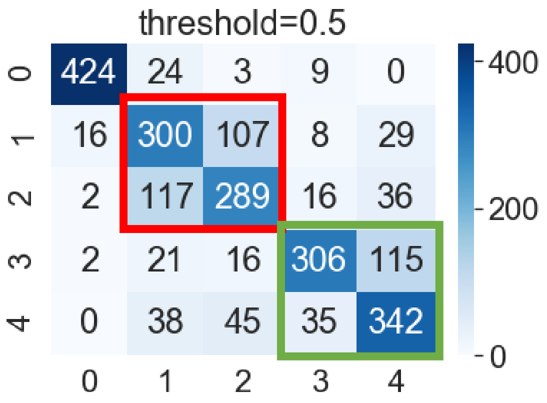
Performance of Multi-class



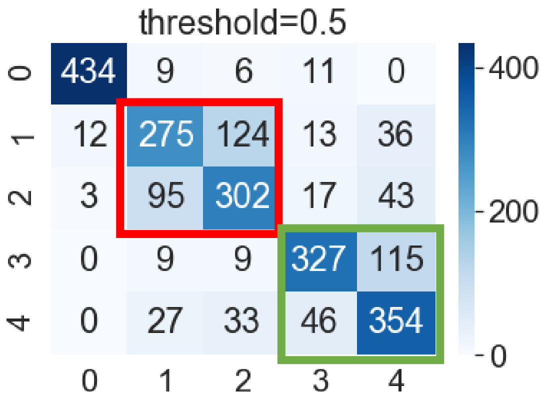
Random Forest



XGBoost



Voting with base models



Voting with all models

Model	Average F1 Score	ROCAUC
Random Forest	0.697	0.916
XGBoost	0.714	0.924
Voting-with-base-models	0.724	0.909
Voting-with-all-models	0.736	0.9338

Voting classifier has largest F1 score and ROC AUC,

Xgboost is the second

Content

Introduction

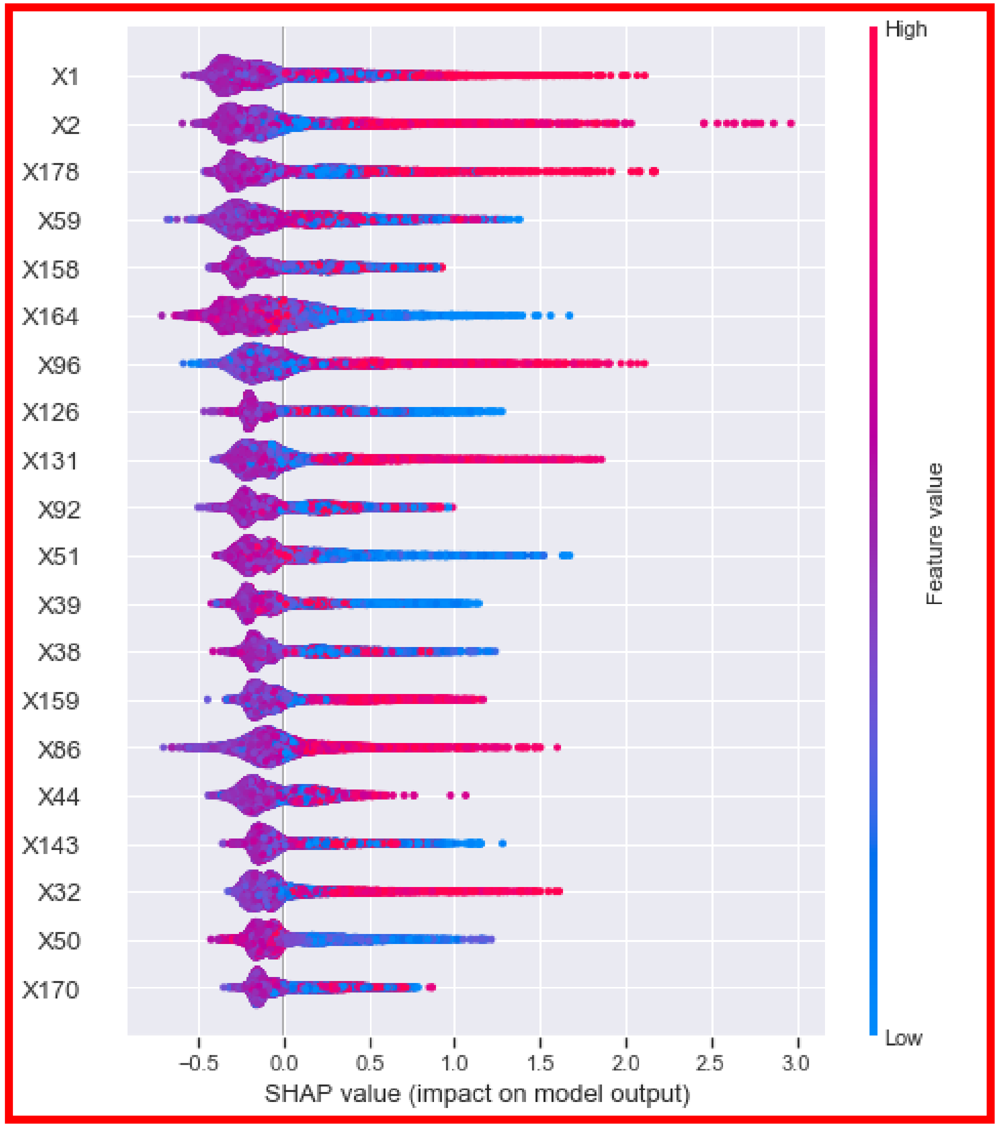
Preprocessing

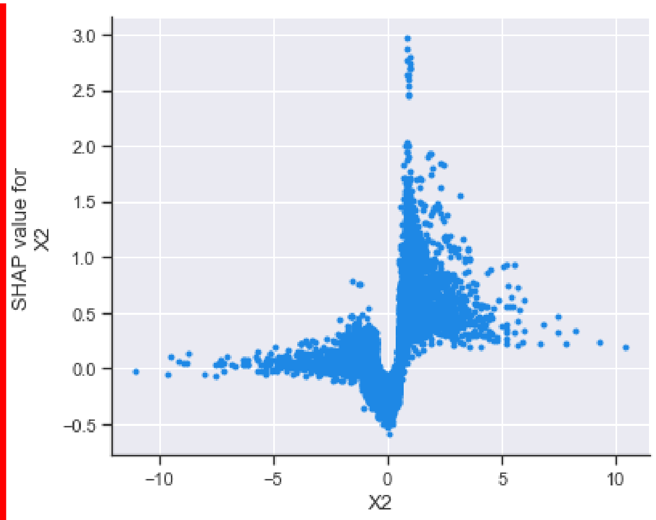
Metrics Selection

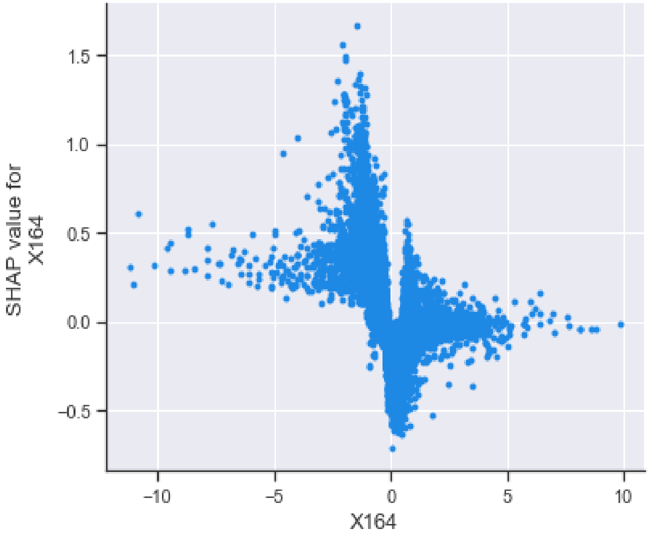
Model evaluation

Summary and Analysis

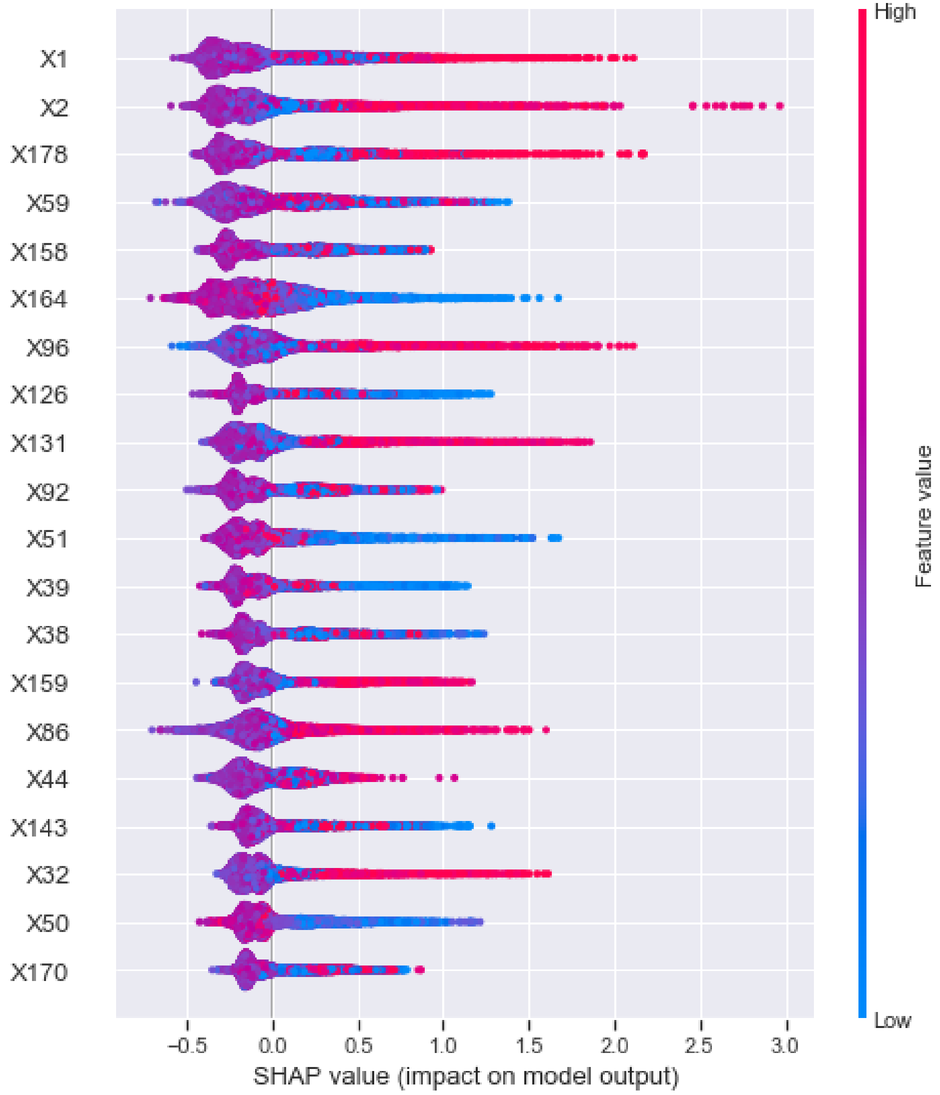
Feature importance in binary classification

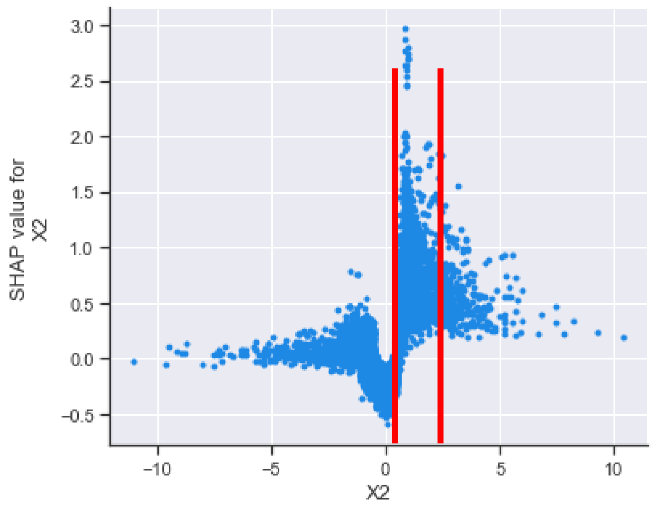






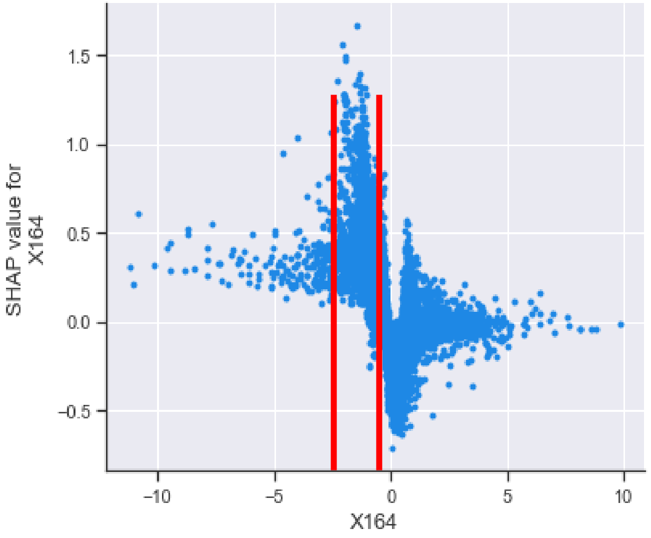
Feature importance in binary classification



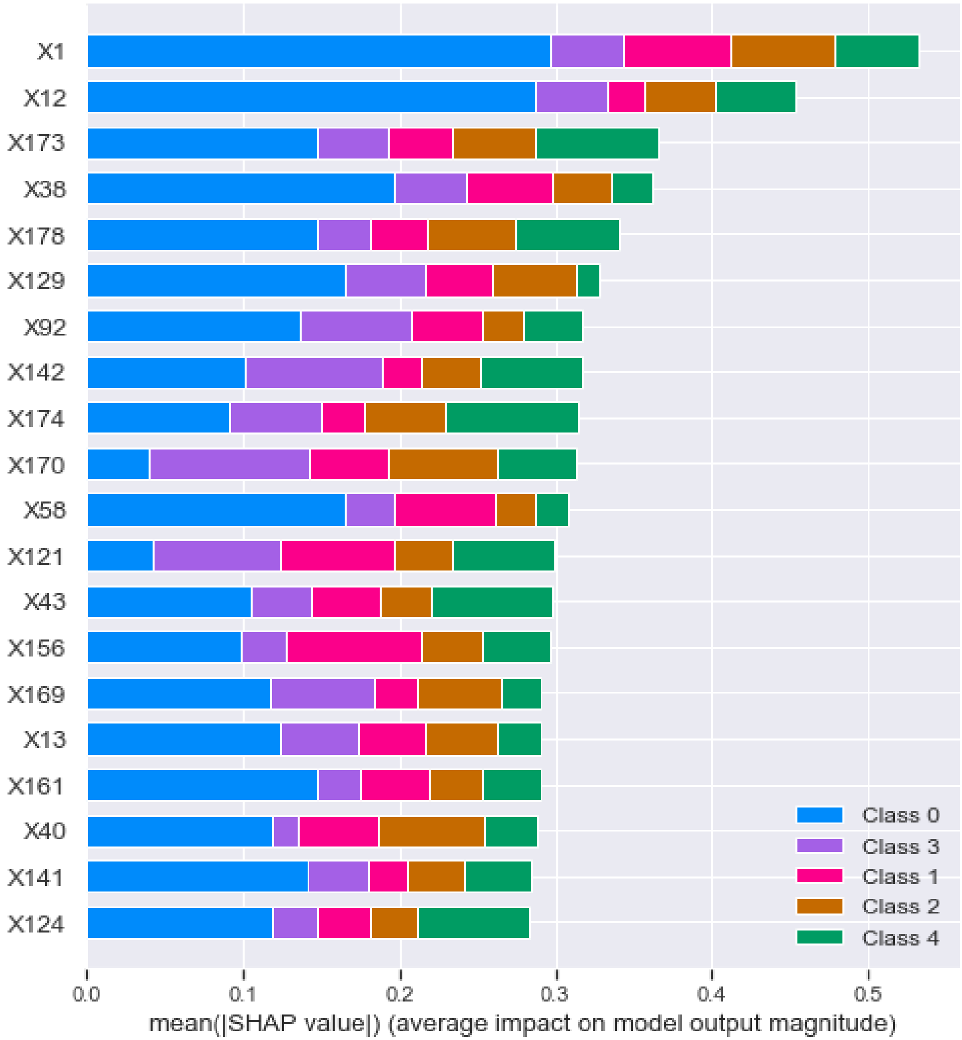


X2 true range,
155.14~404.24,
More likely have seizure

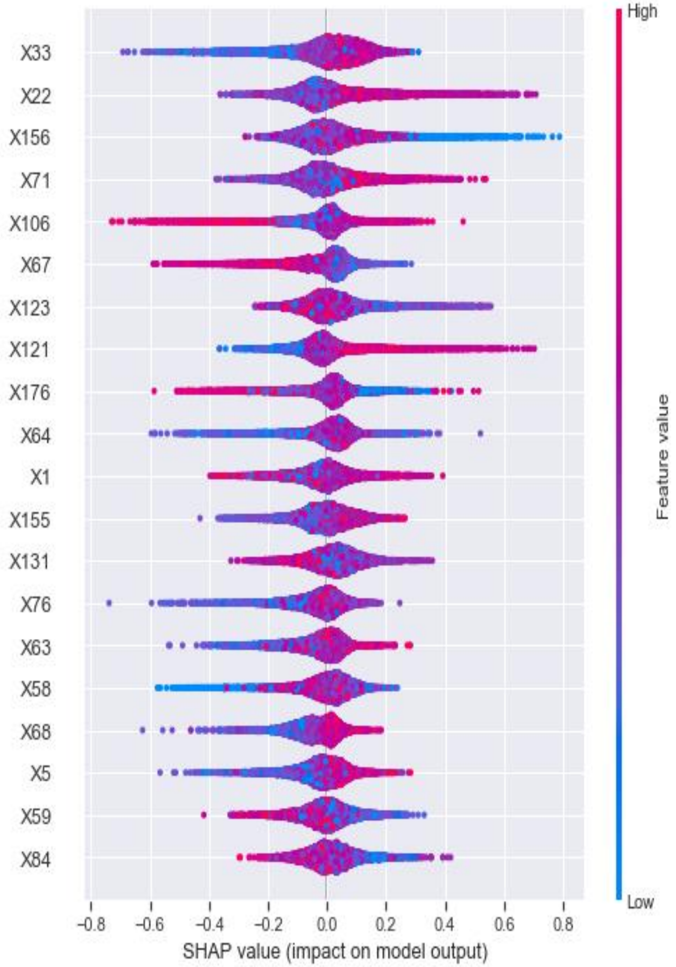
X164 true range,
-427.28~-176.78,
More likely have seizure



Feature importance in multi-classification



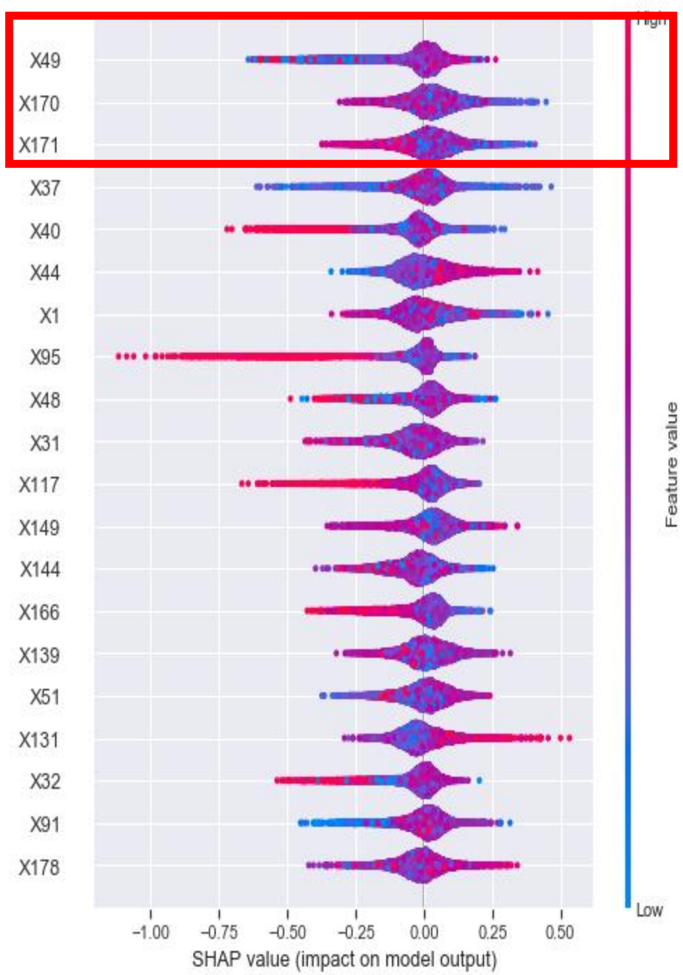
label2



X33, X22, X156

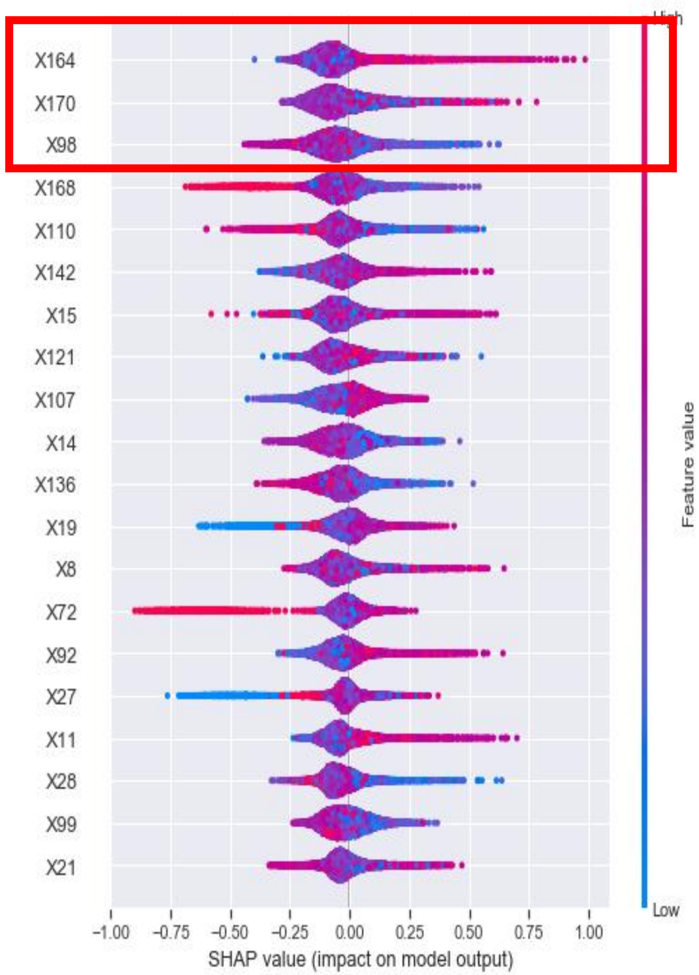
Feature importance in multi-classification

label3



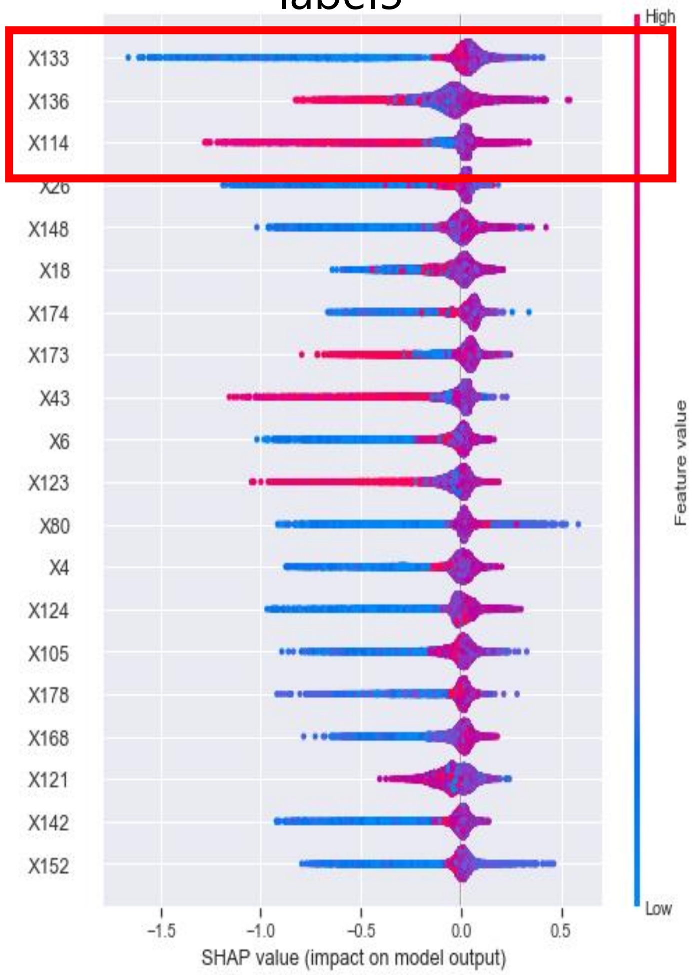
X49, X170 ,X171

label4



X164, X170 ,X98

label5



X133, X136 ,X114.

Summary

Metrics should be chosen carefully, different metrics will lead to choose different model.

voting with all models is the best model in binary and multi-class classification

X1, X2, and X178 is top 3 features for model to predict whether a patient has epileptic seizures, the specific range of features like X2 and X164 will suggest seizure activity

For multi-class classification, X1 X12 and X173 are considered the top 3 important features. Different important features of different class are selected

Limitation and future work

Limitation

- Threshold was selected based on the imbalanced sample. If the unbiased degree of the sample cannot reflect the true distribution of the whole population, we need to re-select our threshold again
- The model of Multi-class classification is used directly from binary classification
- The feature importance about multi-class classification may be not right since the performance of model is not high

Future work

- Select the top20 feature importance and use them to predict and evaluate the model performance
- Fine-tune the model for multi-class classification to generate best model for predicting.
- Find more accurate model to distinguish label 2 and 3 in multi-class classification

Thank you for your listening!

Q&A