Report of training

# Introduction

Report of search and training made on 2021-03-23

## Training data

There are 690 training samples. The distribution of the labels is the following:

* Class 1: 307 instances.
* Class 0: 383 instances.

## Optimizing procedure

The parameters for the bayesian search are:

* Nested Cross Validation using 10 outer folds and 10 inner folds.
* Some of the folds will be skipped. In particular, [0, 2, 4, 6, 8] outer folds and [0, 2, 4, 6, 8] inner folds will be skipped.
* For each outer fold search, a model will be fitted. In order to search for the best hyperparameters, 10 initial points will be evaluated, and 10 additional calls will be made.
* Models will be calibrated using their inner validation set.
* The optimizing metric for the bayesian search is average\_precision.
* The function used for the bayesian search is gp\_minimize.
* Additionally, 20 instances will be left out for assessing the variance of all models.

The search spaces for the optimization is the following:

* Search space for xgboost model.
* model: XGBClassifier(base\_score=None, booster=None, colsample\_bylevel=None,  
   colsample\_bynode=None, colsample\_bytree=None, gamma=None,  
   gpu\_id=None, importance\_type='gain', interaction\_constraints=None,  
   learning\_rate=None, max\_delta\_step=None, max\_depth=None,  
   min\_child\_weight=None, missing=nan, monotone\_constraints=None,  
   n\_estimators=100, n\_jobs=None, num\_parallel\_tree=None,  
   random\_state=None, reg\_alpha=None, reg\_lambda=None,  
   scale\_pos\_weight=None, subsample=None, tree\_method=None,  
   validate\_parameters=None, verbosity=None)
* pipeline\_post\_process: Pipeline(steps=[('post\_process',  
   <nestedcvtraining.utils.pipes\_and\_transformers.OptionedPostProcessTransformer object at 0x000001F401E0EAC0>),  
   ('resample', SMOTE())])
* Search space:
* undersampling\_majority\_class: Categorical(categories=(True, False), prior=None)
* max\_k\_undersampling: Integer(low=5, high=6, prior='uniform', transform='identity')
* resample\_\_sampling\_strategy: Categorical(categories=('minority', 'all'), prior=None)
* post\_process\_\_option: Categorical(categories=('option\_1', 'option\_2', 'option\_3'), prior=None)
* model\_\_max\_depth: Integer(low=5, high=15, prior='uniform', transform='identity')
* model\_\_learning\_rate: Real(low=0.05, high=0.31, prior='log-uniform', transform='identity')
* model\_\_min\_child\_weight: Integer(low=1, high=10, prior='uniform', transform='identity')
* model\_\_subsample: Real(low=0.8, high=1, prior='log-uniform', transform='identity')
* model\_\_colsample\_bytree: Real(low=0.13, high=0.8, prior='log-uniform', transform='identity')
* model\_\_scale\_pos\_weight: Real(low=0.1, high=10, prior='log-uniform', transform='identity')
* model\_\_objective: Categorical(categories=('binary:logistic',), prior=None)
* Search space for random\_forest model.
* model: RandomForestClassifier()
* pipeline\_post\_process: None
* Search space:
* undersampling\_majority\_class: Categorical(categories=(True, False), prior=None)
* model\_\_bootstrap: Integer(low=0, high=1, prior='uniform', transform='identity')
* model\_\_n\_estimators: Integer(low=10, high=100, prior='uniform', transform='identity')
* model\_\_max\_depth: Integer(low=2, high=10, prior='uniform', transform='identity')
* model\_\_min\_samples\_split: Integer(low=5, high=20, prior='uniform', transform='identity')
* model\_\_min\_samples\_leaf: Integer(low=1, high=4, prior='uniform', transform='identity')
* model\_\_max\_features: Categorical(categories=('auto', 'sqrt'), prior=None)
* model\_\_class\_weight: Categorical(categories=('balanced', 'balanced\_subsample'), prior=None)

# Report of validation of the model in the outer Cross Validation

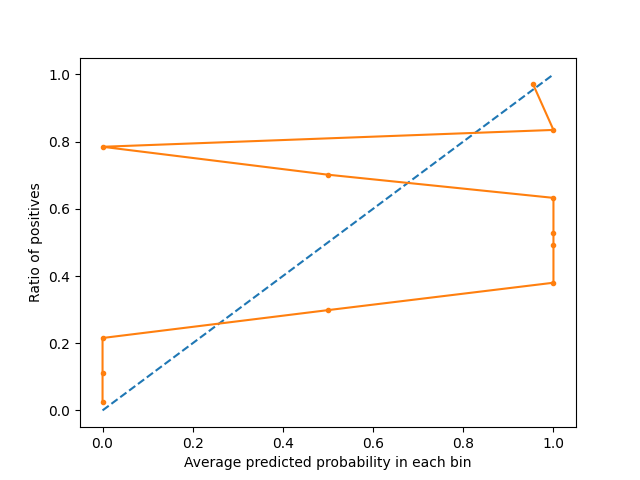
## Winner models of each fold and main metrics

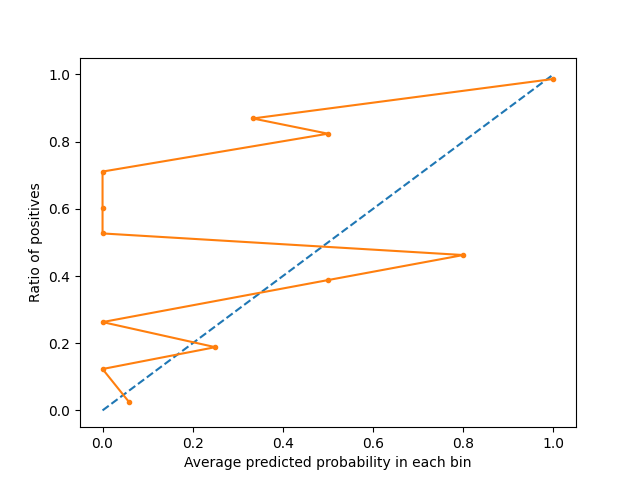
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Fold** | **Model** | **Params** | **Comments** | **roc\_auc** | **neg\_log\_loss** | **average\_precision** | **neg\_brier\_score** |
| 1 | XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_3 * model\_\_max\_depth: 6 * model\_\_learning\_rate: 0.166 * model\_\_min\_child\_weight: 7 * model\_\_subsample: 0.987 * model\_\_colsample\_bytree: 0.228 * model\_\_scale\_pos\_weight: 1.284 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.446 * average size of training set after resampling: 601.2 * average prop of minority class after resampling: 0.5 | 0.979 | 0.228 | 0.972 | 0.069 |
| 3 | RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 47 * model\_\_max\_depth: 7 * model\_\_min\_samples\_split: 6 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.918 | 0.371 | 0.92 | 0.125 |
| 5 | RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 33 * model\_\_max\_depth: 10 * model\_\_min\_samples\_split: 5 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.948 | 0.741 | 0.906 | 0.089 |
| 7 | RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 81 * model\_\_max\_depth: 9 * model\_\_min\_samples\_split: 8 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.964 | 0.264 | 0.96 | 0.079 |
| 9 | XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_3 * model\_\_max\_depth: 7 * model\_\_learning\_rate: 0.177 * model\_\_min\_child\_weight: 6 * model\_\_subsample: 0.971 * model\_\_colsample\_bytree: 0.561 * model\_\_scale\_pos\_weight: 0.156 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.943 | 0.293 | 0.947 | 0.084 |

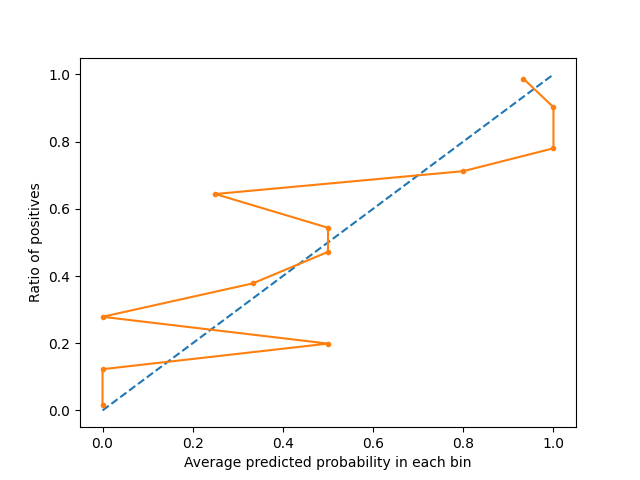
For the selected optimization metric average\_precision the average score is 0.941.

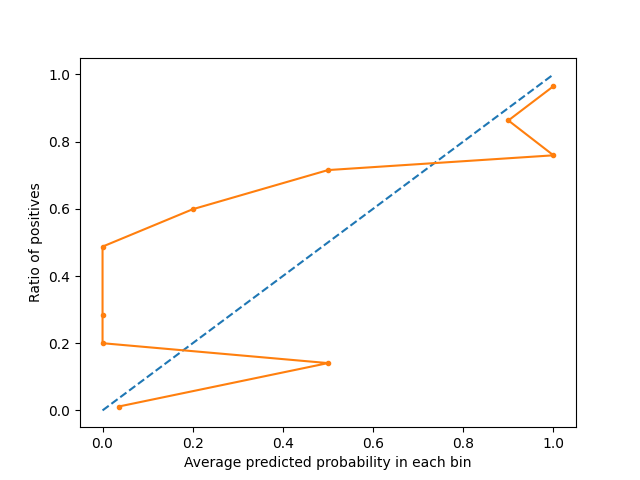
## Main plots

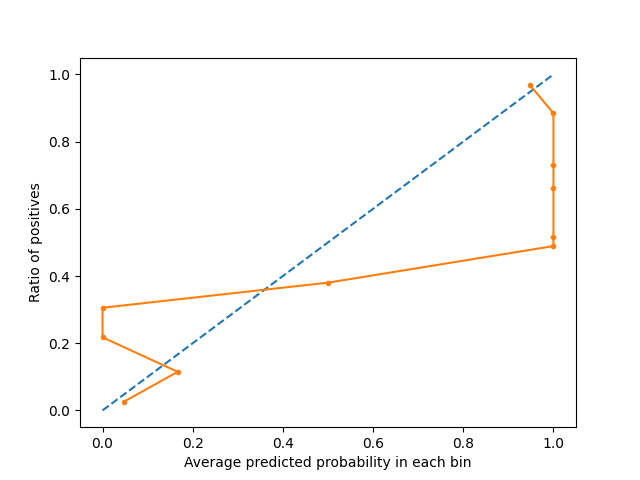
### Calibration plots



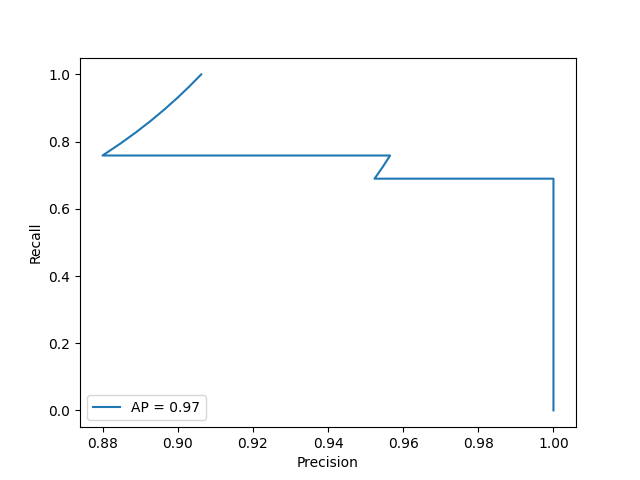


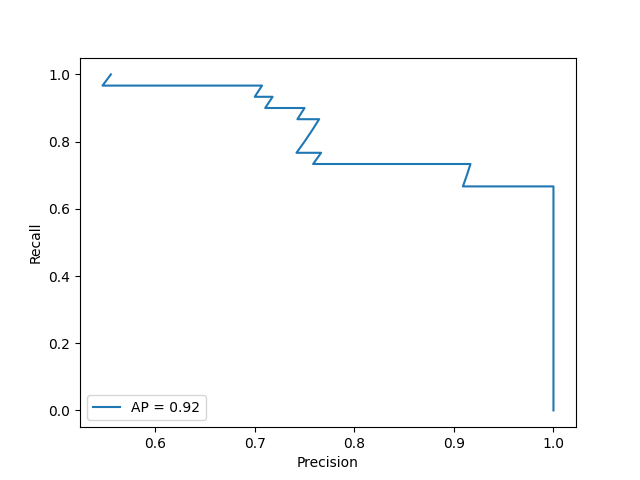


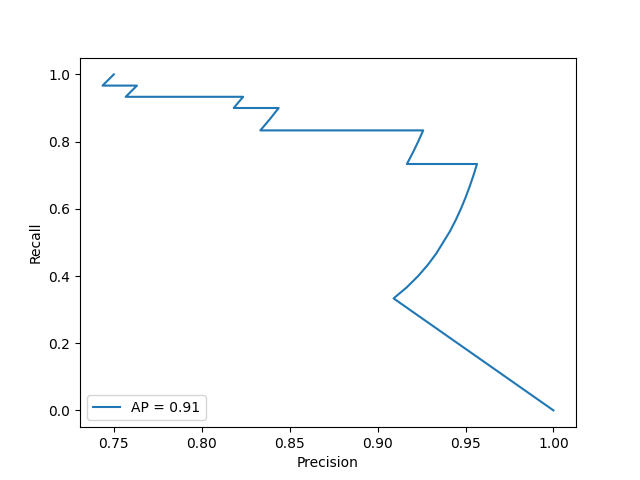


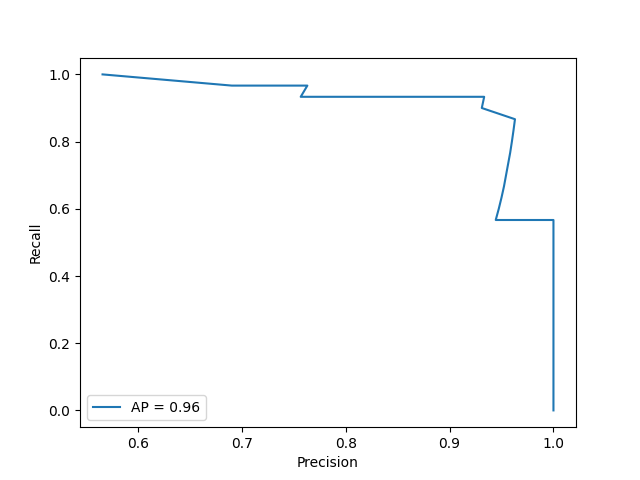


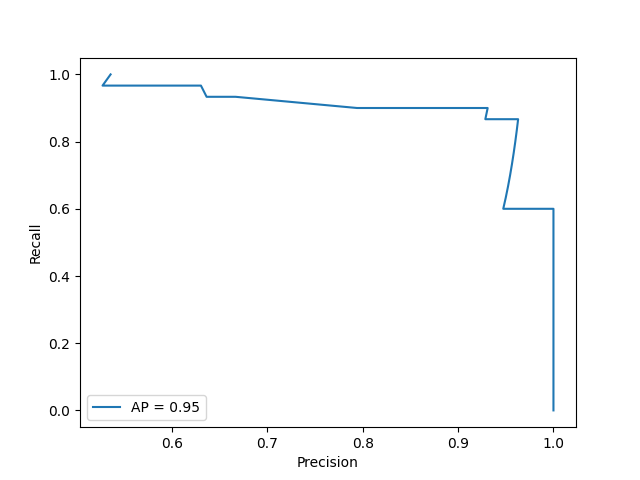
### Precision-recall curve plots



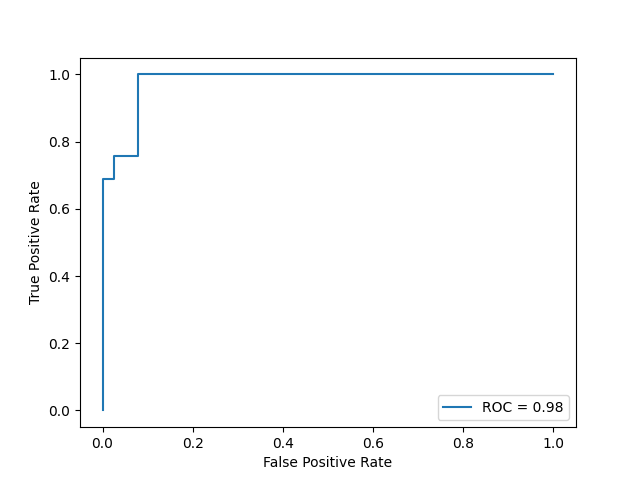


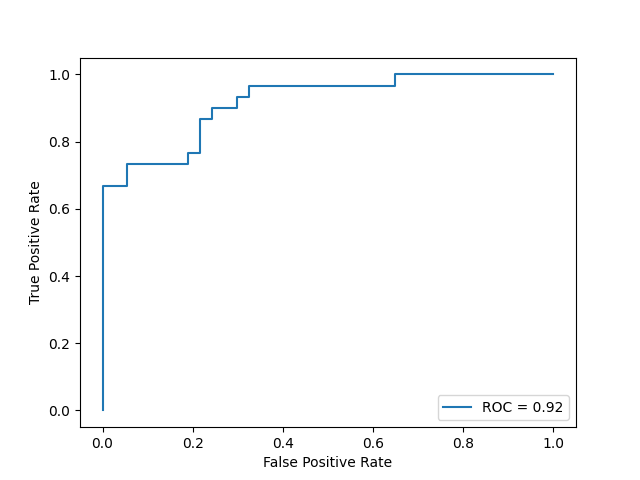


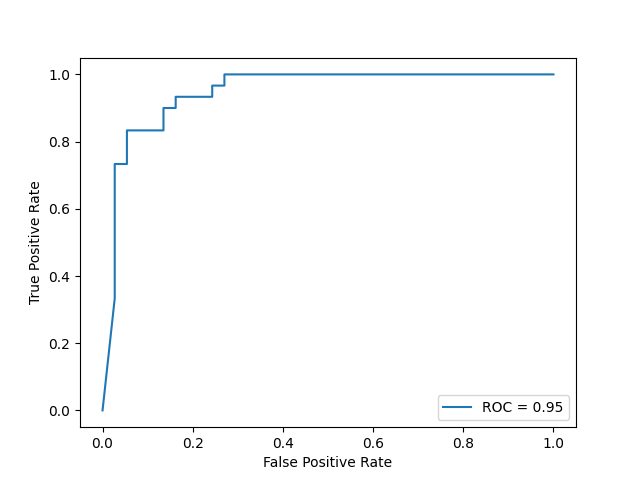


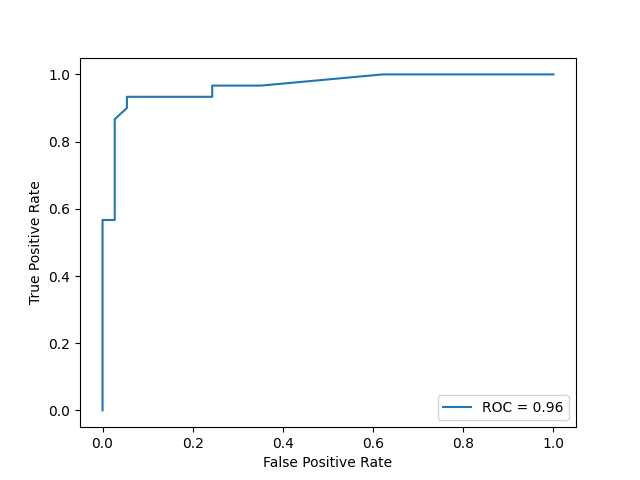


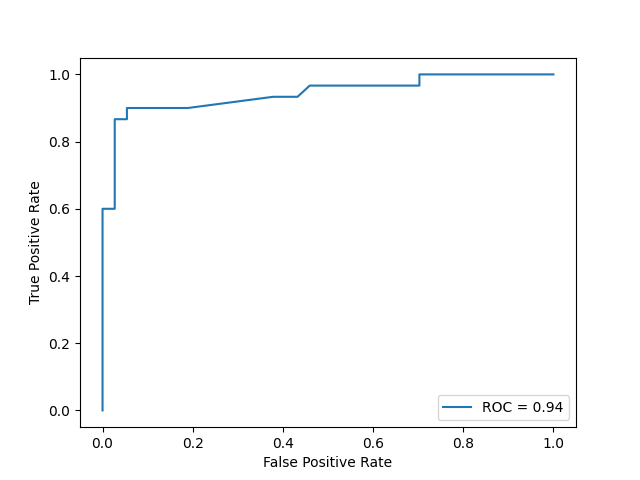
### ROC curve plots



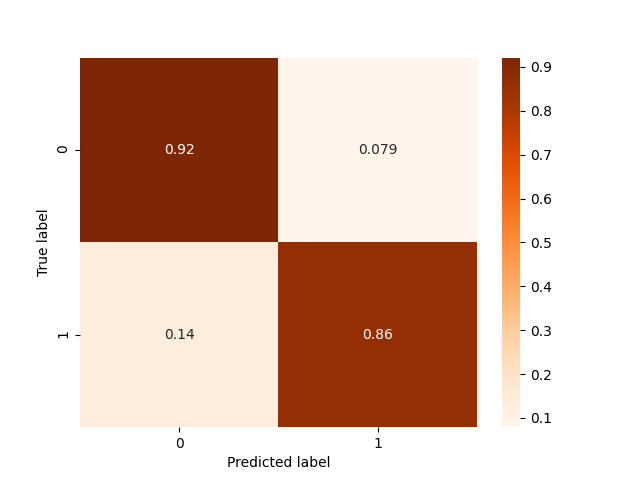


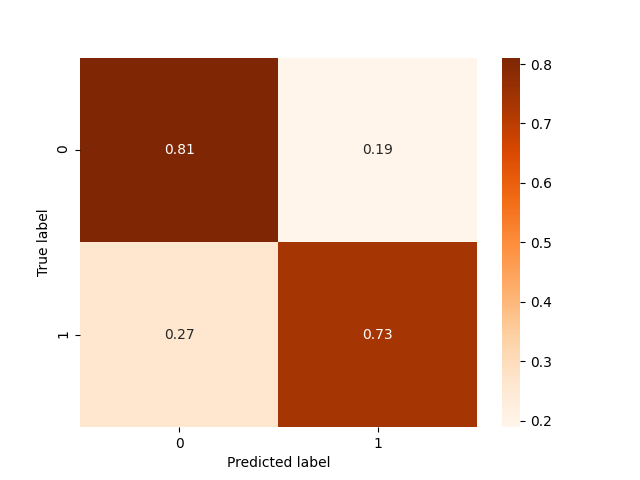


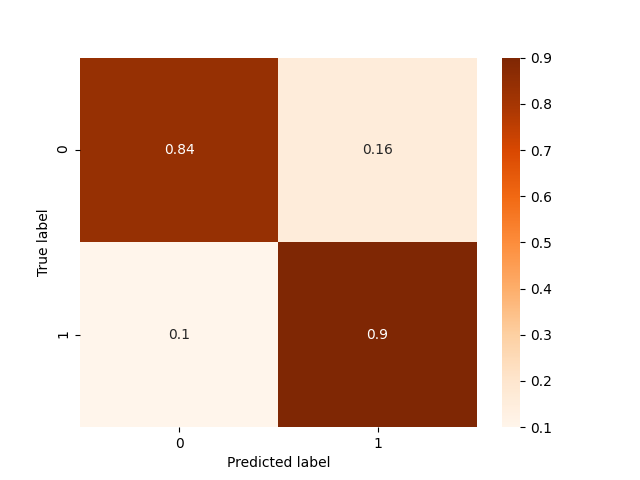


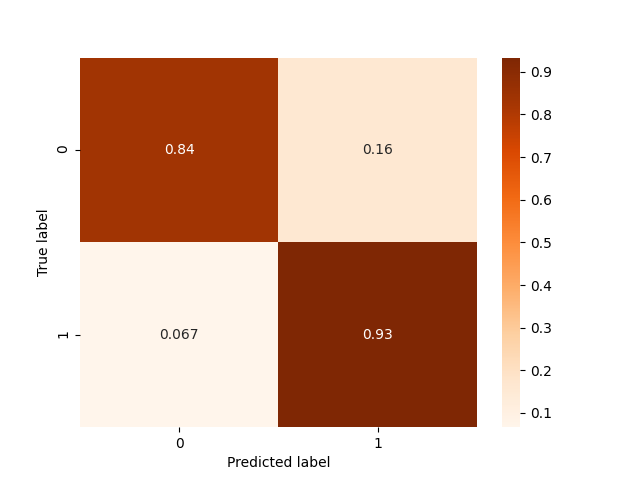


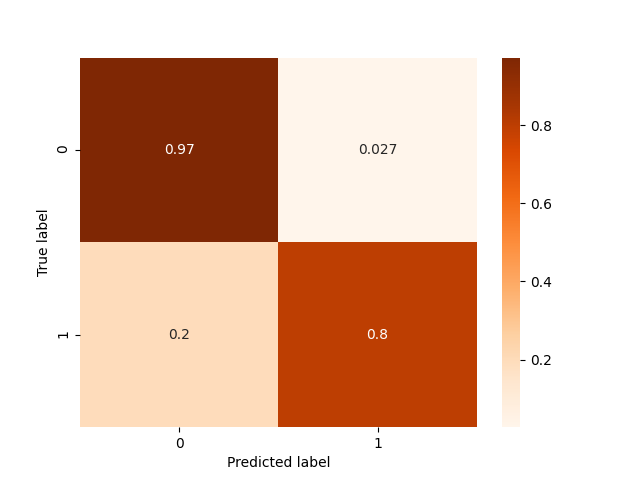
### Confusion matrix



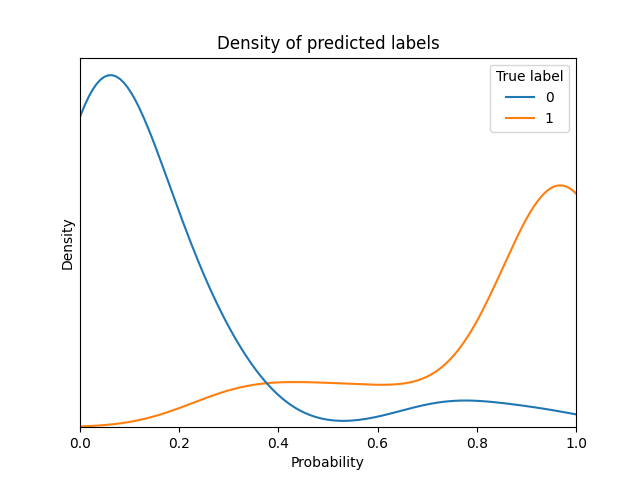


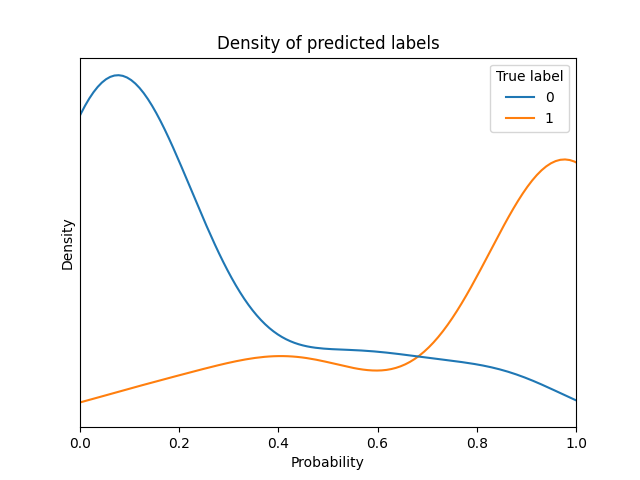


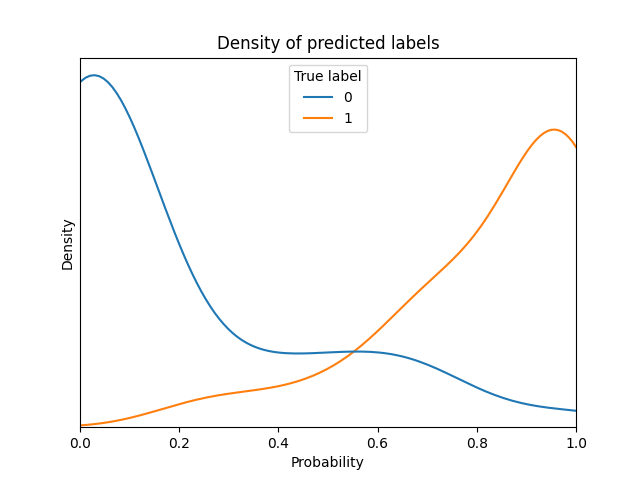


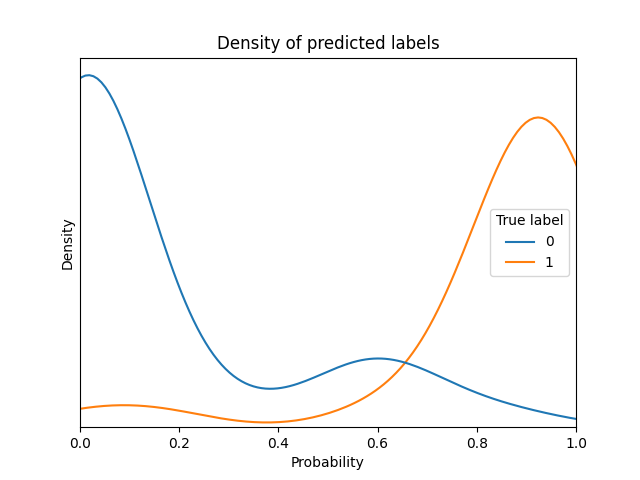


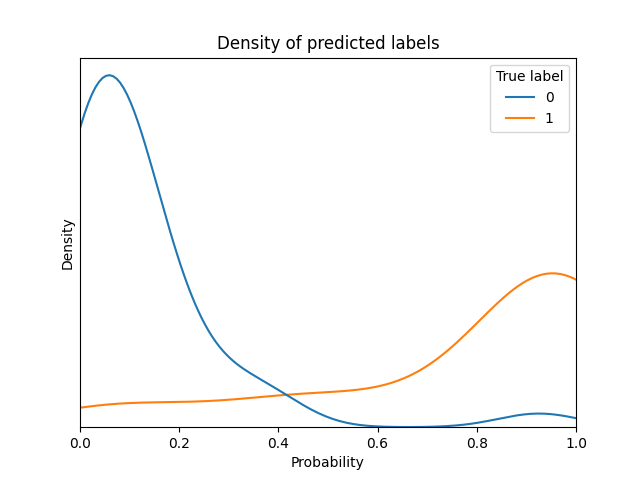
### Histograms











## Comparison of several predictions to assess variance

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Instance** | **Real label** | **Prediction by model of fold 1** | **Prediction by model of fold 3** | **Prediction by model of fold 5** | **Prediction by model of fold 7** | **Prediction by model of fold 9** | **Standard deviation in predictions of this instance** |
| 1 | 0 | 0.013 | 0.017 | 0.179 | 0.014 | 0.041 | 0.064 |
| 2 | 1 | 0.978 | 1.0 | 1.0 | 0.95 | 0.992 | 0.019 |
| 3 | 1 | 0.146 | 0.171 | 0.245 | 0.014 | 0.033 | 0.087 |
| 4 | 0 | 0.033 | 0.187 | 0.245 | 0.169 | 0.103 | 0.073 |
| 5 | 0 | 0.403 | 0.171 | 0.245 | 0.078 | 0.389 | 0.125 |
| 6 | 0 | 0.084 | 0.218 | 0.054 | 0.197 | 0.054 | 0.072 |
| 7 | 1 | 1.0 | 0.995 | 0.986 | 0.854 | 0.958 | 0.054 |
| 8 | 1 | 1.0 | 1.0 | 1.0 | 0.99 | 0.992 | 0.004 |
| 9 | 1 | 1.0 | 1.0 | 1.0 | 0.99 | 0.958 | 0.016 |
| 10 | 0 | 0.137 | 0.017 | 0.067 | 0.014 | 0.302 | 0.107 |
| 11 | 0 | 0.137 | 0.017 | 0.0 | 0.0 | 0.126 | 0.062 |
| 12 | 1 | 1.0 | 1.0 | 1.0 | 0.99 | 0.992 | 0.004 |
| 13 | 1 | 1.0 | 1.0 | 1.0 | 0.95 | 1.0 | 0.02 |
| 14 | 0 | 0.335 | 0.204 | 0.117 | 0.128 | 0.087 | 0.089 |
| 15 | 0 | 0.497 | 0.507 | 0.401 | 0.548 | 0.467 | 0.049 |
| 16 | 0 | 0.102 | 0.413 | 0.332 | 0.375 | 0.126 | 0.13 |
| 17 | 0 | 0.079 | 0.23 | 0.312 | 0.169 | 0.126 | 0.081 |
| 18 | 1 | 0.435 | 0.571 | 0.626 | 0.6 | 0.73 | 0.095 |
| 19 | 0 | 0.574 | 0.915 | 0.749 | 0.853 | 0.73 | 0.117 |
| 20 | 1 | 1.0 | 1.0 | 1.0 | 0.99 | 0.992 | 0.004 |

The average standard deviation is 0.064

# Report of inner trainings

## Report of inner training in fold 1 of outer Cross Validation

## Report of training in this outer fold

Best model with respect to selected metric is XGBClassifier with the following params:

* undersampling\_majority\_class: False
* max\_k\_undersampling: 6
* resample\_\_sampling\_strategy: all
* post\_process\_\_option: option\_3
* model\_\_max\_depth: 6
* model\_\_learning\_rate: 0.166
* model\_\_min\_child\_weight: 7
* model\_\_subsample: 0.987
* model\_\_colsample\_bytree: 0.228
* model\_\_scale\_pos\_weight: 1.284
* model\_\_objective: binary:logistic

### Comparison of all models trained in this outer fold

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Params** | **Comments** | **roc\_auc on inner fold** | **neg\_log\_loss on inner fold** | **average\_precision on inner fold** | **neg\_brier\_score on inner fold** | **roc\_auc on outer fold** | **neg\_log\_loss on outer fold** | **average\_precision on outer fold** | **neg\_brier\_score on outer fold** |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_1 * model\_\_max\_depth: 10 * model\_\_learning\_rate: 0.216 * model\_\_min\_child\_weight: 7 * model\_\_subsample: 0.869 * model\_\_colsample\_bytree: 0.177 * model\_\_scale\_pos\_weight: 1.337 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.446 * average size of training set after resampling: 601.2 * average prop of minority class after resampling: 0.5 | 0.917 | 0.314 | 0.876 | 0.106 | 0.885 | 0.895 | 0.883 | 0.133 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_3 * model\_\_max\_depth: 6 * model\_\_learning\_rate: 0.166 * model\_\_min\_child\_weight: 7 * model\_\_subsample: 0.987 * model\_\_colsample\_bytree: 0.228 * model\_\_scale\_pos\_weight: 1.284 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.446 * average size of training set after resampling: 601.2 * average prop of minority class after resampling: 0.5 | 0.968 | 0.191 | 0.953 | 0.062 | 0.979 | 0.228 | 0.972 | 0.069 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 15 * model\_\_learning\_rate: 0.135 * model\_\_min\_child\_weight: 7 * model\_\_subsample: 0.821 * model\_\_colsample\_bytree: 0.656 * model\_\_scale\_pos\_weight: 0.933 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.446 * average size of training set after resampling: 601.2 * average prop of minority class after resampling: 0.5 | 0.945 | 0.247 | 0.925 | 0.08 | 0.956 | 0.283 | 0.945 | 0.088 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 12 * model\_\_learning\_rate: 0.158 * model\_\_min\_child\_weight: 5 * model\_\_subsample: 0.963 * model\_\_colsample\_bytree: 0.192 * model\_\_scale\_pos\_weight: 0.812 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.5 * average prop of minority class before resampling: 0.383 * average size of training set after resampling: 484.4 * average prop of minority class after resampling: 0.5 | 0.952 | 0.231 | 0.939 | 0.074 | 0.965 | 0.254 | 0.959 | 0.08 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 8 * model\_\_learning\_rate: 0.062 * model\_\_min\_child\_weight: 4 * model\_\_subsample: 0.801 * model\_\_colsample\_bytree: 0.286 * model\_\_scale\_pos\_weight: 4.723 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.446 * average size of training set after resampling: 601.2 * average prop of minority class after resampling: 0.5 | 0.952 | 0.235 | 0.933 | 0.077 | 0.946 | 0.308 | 0.944 | 0.091 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_1 * model\_\_max\_depth: 9 * model\_\_learning\_rate: 0.061 * model\_\_min\_child\_weight: 3 * model\_\_subsample: 0.911 * model\_\_colsample\_bytree: 0.294 * model\_\_scale\_pos\_weight: 0.104 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.5 * average prop of minority class before resampling: 0.383 * average size of training set after resampling: 484.4 * average prop of minority class after resampling: 0.5 | 0.939 | 0.274 | 0.909 | 0.088 | 0.941 | 0.318 | 0.932 | 0.098 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_3 * model\_\_max\_depth: 6 * model\_\_learning\_rate: 0.134 * model\_\_min\_child\_weight: 4 * model\_\_subsample: 0.994 * model\_\_colsample\_bytree: 0.28 * model\_\_scale\_pos\_weight: 0.155 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.446 * average size of training set after resampling: 601.2 * average prop of minority class after resampling: 0.5 | 0.957 | 0.212 | 0.941 | 0.067 | 0.963 | 0.252 | 0.953 | 0.076 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 14 * model\_\_learning\_rate: 0.202 * model\_\_min\_child\_weight: 3 * model\_\_subsample: 0.913 * model\_\_colsample\_bytree: 0.149 * model\_\_scale\_pos\_weight: 1.507 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.5 * average prop of minority class before resampling: 0.383 * average size of training set after resampling: 484.4 * average prop of minority class after resampling: 0.5 | 0.946 | 0.249 | 0.925 | 0.081 | 0.945 | 0.296 | 0.94 | 0.088 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 13 * model\_\_learning\_rate: 0.056 * model\_\_min\_child\_weight: 6 * model\_\_subsample: 0.81 * model\_\_colsample\_bytree: 0.274 * model\_\_scale\_pos\_weight: 0.421 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.5 * average prop of minority class before resampling: 0.383 * average size of training set after resampling: 484.4 * average prop of minority class after resampling: 0.5 | 0.951 | 0.231 | 0.934 | 0.073 | 0.946 | 0.296 | 0.937 | 0.089 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_1 * model\_\_max\_depth: 11 * model\_\_learning\_rate: 0.254 * model\_\_min\_child\_weight: 4 * model\_\_subsample: 0.968 * model\_\_colsample\_bytree: 0.181 * model\_\_scale\_pos\_weight: 9.124 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.5 * average prop of minority class before resampling: 0.383 * average size of training set after resampling: 484.4 * average prop of minority class after resampling: 0.5 | 0.909 | 0.335 | 0.868 | 0.111 | 0.891 | 0.406 | 0.895 | 0.121 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 45 * model\_\_max\_depth: 5 * model\_\_min\_samples\_split: 16 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.5 * average prop of minority class: 0.383 | 0.964 | 0.195 | 0.943 | 0.064 | 0.965 | 0.252 | 0.947 | 0.072 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 1 * model\_\_n\_estimators: 88 * model\_\_max\_depth: 8 * model\_\_min\_samples\_split: 16 * model\_\_min\_samples\_leaf: 1 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.446 | 0.964 | 0.191 | 0.947 | 0.061 | 0.953 | 0.725 | 0.909 | 0.078 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 1 * model\_\_n\_estimators: 34 * model\_\_max\_depth: 5 * model\_\_min\_samples\_split: 19 * model\_\_min\_samples\_leaf: 1 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.5 * average prop of minority class: 0.383 | 0.962 | 0.199 | 0.942 | 0.064 | 0.958 | 0.267 | 0.935 | 0.075 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 43 * model\_\_max\_depth: 7 * model\_\_min\_samples\_split: 5 * model\_\_min\_samples\_leaf: 1 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.446 | 0.963 | 0.195 | 0.943 | 0.063 | 0.955 | 0.276 | 0.933 | 0.08 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 1 * model\_\_n\_estimators: 53 * model\_\_max\_depth: 3 * model\_\_min\_samples\_split: 8 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: auto * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.446 | 0.958 | 0.217 | 0.941 | 0.07 | 0.941 | 0.73 | 0.909 | 0.077 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 91 * model\_\_max\_depth: 5 * model\_\_min\_samples\_split: 8 * model\_\_min\_samples\_leaf: 4 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 392.5 * average prop of minority class: 0.383 | 0.963 | 0.198 | 0.947 | 0.065 | 0.956 | 0.717 | 0.92 | 0.077 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 1 * model\_\_n\_estimators: 91 * model\_\_max\_depth: 4 * model\_\_min\_samples\_split: 6 * model\_\_min\_samples\_leaf: 4 * model\_\_max\_features: auto * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 392.5 * average prop of minority class: 0.383 | 0.96 | 0.213 | 0.941 | 0.07 | 0.953 | 0.714 | 0.909 | 0.074 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 83 * model\_\_max\_depth: 6 * model\_\_min\_samples\_split: 11 * model\_\_min\_samples\_leaf: 4 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.446 | 0.963 | 0.195 | 0.942 | 0.064 | 0.96 | 0.255 | 0.945 | 0.075 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 37 * model\_\_max\_depth: 5 * model\_\_min\_samples\_split: 12 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.5 * average prop of minority class: 0.383 | 0.965 | 0.19 | 0.945 | 0.063 | 0.952 | 0.719 | 0.915 | 0.077 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 1 * model\_\_n\_estimators: 98 * model\_\_max\_depth: 5 * model\_\_min\_samples\_split: 19 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: auto * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 392.5 * average prop of minority class: 0.383 | 0.961 | 0.205 | 0.939 | 0.066 | 0.951 | 0.716 | 0.913 | 0.074 |

## Report of inner training in fold 3 of outer Cross Validation

## Report of training in this outer fold

Best model with respect to selected metric is RandomForestClassifier with the following params:

* undersampling\_majority\_class: False
* model\_\_bootstrap: 0
* model\_\_n\_estimators: 47
* model\_\_max\_depth: 7
* model\_\_min\_samples\_split: 6
* model\_\_min\_samples\_leaf: 2
* model\_\_max\_features: auto
* model\_\_class\_weight: balanced\_subsample

### Comparison of all models trained in this outer fold

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Params** | **Comments** | **roc\_auc on inner fold** | **neg\_log\_loss on inner fold** | **average\_precision on inner fold** | **neg\_brier\_score on inner fold** | **roc\_auc on outer fold** | **neg\_log\_loss on outer fold** | **average\_precision on outer fold** | **neg\_brier\_score on outer fold** |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 6 * model\_\_learning\_rate: 0.083 * model\_\_min\_child\_weight: 10 * model\_\_subsample: 0.824 * model\_\_colsample\_bytree: 0.156 * model\_\_scale\_pos\_weight: 0.228 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.946 | 0.261 | 0.918 | 0.081 | 0.904 | 0.409 | 0.901 | 0.128 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 5 * model\_\_learning\_rate: 0.29 * model\_\_min\_child\_weight: 5 * model\_\_subsample: 1.0 * model\_\_colsample\_bytree: 0.782 * model\_\_scale\_pos\_weight: 0.222 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.954 | 0.233 | 0.93 | 0.074 | 0.884 | 0.474 | 0.884 | 0.15 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_3 * model\_\_max\_depth: 14 * model\_\_learning\_rate: 0.053 * model\_\_min\_child\_weight: 4 * model\_\_subsample: 0.872 * model\_\_colsample\_bytree: 0.741 * model\_\_scale\_pos\_weight: 0.932 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.969 | 0.178 | 0.951 | 0.056 | 0.922 | 0.363 | 0.927 | 0.12 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 13 * model\_\_learning\_rate: 0.202 * model\_\_min\_child\_weight: 9 * model\_\_subsample: 0.952 * model\_\_colsample\_bytree: 0.436 * model\_\_scale\_pos\_weight: 8.802 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.949 | 0.246 | 0.93 | 0.077 | 0.884 | 0.483 | 0.893 | 0.16 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_3 * model\_\_max\_depth: 10 * model\_\_learning\_rate: 0.225 * model\_\_min\_child\_weight: 3 * model\_\_subsample: 0.824 * model\_\_colsample\_bytree: 0.548 * model\_\_scale\_pos\_weight: 0.136 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.953 | 0.233 | 0.93 | 0.072 | 0.906 | 0.414 | 0.91 | 0.134 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 7 * model\_\_learning\_rate: 0.069 * model\_\_min\_child\_weight: 6 * model\_\_subsample: 0.844 * model\_\_colsample\_bytree: 0.238 * model\_\_scale\_pos\_weight: 0.146 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.956 | 0.228 | 0.938 | 0.07 | 0.905 | 0.446 | 0.904 | 0.139 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 15 * model\_\_learning\_rate: 0.091 * model\_\_min\_child\_weight: 4 * model\_\_subsample: 0.977 * model\_\_colsample\_bytree: 0.504 * model\_\_scale\_pos\_weight: 0.306 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.951 | 0.24 | 0.93 | 0.077 | 0.907 | 0.389 | 0.904 | 0.133 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 15 * model\_\_learning\_rate: 0.291 * model\_\_min\_child\_weight: 3 * model\_\_subsample: 0.834 * model\_\_colsample\_bytree: 0.476 * model\_\_scale\_pos\_weight: 1.933 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.943 | 0.262 | 0.919 | 0.085 | 0.891 | 0.433 | 0.895 | 0.148 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 14 * model\_\_learning\_rate: 0.067 * model\_\_min\_child\_weight: 7 * model\_\_subsample: 0.946 * model\_\_colsample\_bytree: 0.223 * model\_\_scale\_pos\_weight: 1.118 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.958 | 0.22 | 0.942 | 0.069 | 0.898 | 0.451 | 0.901 | 0.147 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 8 * model\_\_learning\_rate: 0.094 * model\_\_min\_child\_weight: 10 * model\_\_subsample: 0.994 * model\_\_colsample\_bytree: 0.794 * model\_\_scale\_pos\_weight: 0.179 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.954 | 0.237 | 0.932 | 0.075 | 0.907 | 0.421 | 0.911 | 0.136 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 47 * model\_\_max\_depth: 3 * model\_\_min\_samples\_split: 11 * model\_\_min\_samples\_leaf: 1 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.958 | 0.229 | 0.937 | 0.071 | 0.901 | 0.41 | 0.899 | 0.134 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 43 * model\_\_max\_depth: 9 * model\_\_min\_samples\_split: 11 * model\_\_min\_samples\_leaf: 1 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.966 | 0.195 | 0.946 | 0.064 | 0.93 | 0.341 | 0.928 | 0.113 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 30 * model\_\_max\_depth: 9 * model\_\_min\_samples\_split: 10 * model\_\_min\_samples\_leaf: 1 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.969 | 0.181 | 0.951 | 0.058 | 0.92 | 0.371 | 0.919 | 0.126 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 49 * model\_\_max\_depth: 10 * model\_\_min\_samples\_split: 13 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.968 | 0.188 | 0.949 | 0.06 | 0.906 | 0.435 | 0.905 | 0.14 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 93 * model\_\_max\_depth: 6 * model\_\_min\_samples\_split: 11 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.965 | 0.194 | 0.945 | 0.063 | 0.912 | 0.385 | 0.909 | 0.129 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 47 * model\_\_max\_depth: 7 * model\_\_min\_samples\_split: 6 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.969 | 0.183 | 0.953 | 0.059 | 0.918 | 0.371 | 0.92 | 0.125 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 84 * model\_\_max\_depth: 8 * model\_\_min\_samples\_split: 17 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.966 | 0.195 | 0.946 | 0.063 | 0.921 | 0.372 | 0.918 | 0.125 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 1 * model\_\_n\_estimators: 51 * model\_\_max\_depth: 3 * model\_\_min\_samples\_split: 14 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.965 | 0.201 | 0.943 | 0.063 | 0.909 | 0.418 | 0.905 | 0.131 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 80 * model\_\_max\_depth: 7 * model\_\_min\_samples\_split: 16 * model\_\_min\_samples\_leaf: 4 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.969 | 0.179 | 0.949 | 0.057 | 0.914 | 0.403 | 0.915 | 0.13 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 25 * model\_\_max\_depth: 6 * model\_\_min\_samples\_split: 18 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: auto * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.962 | 0.208 | 0.94 | 0.067 | 0.908 | 0.406 | 0.907 | 0.134 |

## Report of inner training in fold 5 of outer Cross Validation

## Report of training in this outer fold

Best model with respect to selected metric is RandomForestClassifier with the following params:

* undersampling\_majority\_class: False
* model\_\_bootstrap: 0
* model\_\_n\_estimators: 33
* model\_\_max\_depth: 10
* model\_\_min\_samples\_split: 5
* model\_\_min\_samples\_leaf: 3
* model\_\_max\_features: sqrt
* model\_\_class\_weight: balanced

### Comparison of all models trained in this outer fold

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Params** | **Comments** | **roc\_auc on inner fold** | **neg\_log\_loss on inner fold** | **average\_precision on inner fold** | **neg\_brier\_score on inner fold** | **roc\_auc on outer fold** | **neg\_log\_loss on outer fold** | **average\_precision on outer fold** | **neg\_brier\_score on outer fold** |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_1 * model\_\_max\_depth: 14 * model\_\_learning\_rate: 0.075 * model\_\_min\_child\_weight: 3 * model\_\_subsample: 0.878 * model\_\_colsample\_bytree: 0.196 * model\_\_scale\_pos\_weight: 1.263 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.912 | 0.339 | 0.869 | 0.111 | 0.924 | 0.329 | 0.926 | 0.097 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 12 * model\_\_learning\_rate: 0.302 * model\_\_min\_child\_weight: 3 * model\_\_subsample: 0.877 * model\_\_colsample\_bytree: 0.158 * model\_\_scale\_pos\_weight: 3.823 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.945 | 0.246 | 0.906 | 0.081 | 0.948 | 0.282 | 0.937 | 0.087 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_3 * model\_\_max\_depth: 7 * model\_\_learning\_rate: 0.306 * model\_\_min\_child\_weight: 2 * model\_\_subsample: 0.878 * model\_\_colsample\_bytree: 0.348 * model\_\_scale\_pos\_weight: 0.802 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.967 | 0.186 | 0.942 | 0.06 | 0.953 | 0.28 | 0.946 | 0.09 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_1 * model\_\_max\_depth: 8 * model\_\_learning\_rate: 0.236 * model\_\_min\_child\_weight: 6 * model\_\_subsample: 0.895 * model\_\_colsample\_bytree: 0.21 * model\_\_scale\_pos\_weight: 4.994 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.91 | 0.339 | 0.864 | 0.113 | 0.901 | 0.362 | 0.92 | 0.11 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 14 * model\_\_learning\_rate: 0.177 * model\_\_min\_child\_weight: 7 * model\_\_subsample: 0.836 * model\_\_colsample\_bytree: 0.287 * model\_\_scale\_pos\_weight: 0.106 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.935 | 0.278 | 0.904 | 0.091 | 0.95 | 0.295 | 0.938 | 0.094 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_1 * model\_\_max\_depth: 6 * model\_\_learning\_rate: 0.084 * model\_\_min\_child\_weight: 2 * model\_\_subsample: 0.828 * model\_\_colsample\_bytree: 0.2 * model\_\_scale\_pos\_weight: 1.034 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.9 | 0.357 | 0.856 | 0.117 | 0.908 | 0.386 | 0.899 | 0.121 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 12 * model\_\_learning\_rate: 0.153 * model\_\_min\_child\_weight: 4 * model\_\_subsample: 0.857 * model\_\_colsample\_bytree: 0.431 * model\_\_scale\_pos\_weight: 7.419 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.942 | 0.257 | 0.907 | 0.085 | 0.943 | 0.291 | 0.931 | 0.088 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 11 * model\_\_learning\_rate: 0.096 * model\_\_min\_child\_weight: 9 * model\_\_subsample: 0.868 * model\_\_colsample\_bytree: 0.632 * model\_\_scale\_pos\_weight: 2.333 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.951 | 0.236 | 0.917 | 0.076 | 0.951 | 0.27 | 0.933 | 0.088 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_3 * model\_\_max\_depth: 8 * model\_\_learning\_rate: 0.055 * model\_\_min\_child\_weight: 9 * model\_\_subsample: 0.863 * model\_\_colsample\_bytree: 0.475 * model\_\_scale\_pos\_weight: 0.417 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.956 | 0.227 | 0.931 | 0.07 | 0.961 | 0.258 | 0.953 | 0.081 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 12 * model\_\_learning\_rate: 0.248 * model\_\_min\_child\_weight: 2 * model\_\_subsample: 0.938 * model\_\_colsample\_bytree: 0.282 * model\_\_scale\_pos\_weight: 0.202 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.949 | 0.238 | 0.917 | 0.077 | 0.952 | 0.305 | 0.947 | 0.093 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 1 * model\_\_n\_estimators: 53 * model\_\_max\_depth: 4 * model\_\_min\_samples\_split: 17 * model\_\_min\_samples\_leaf: 4 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.961 | 0.213 | 0.932 | 0.068 | 0.954 | 0.265 | 0.942 | 0.087 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 52 * model\_\_max\_depth: 2 * model\_\_min\_samples\_split: 18 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.947 | 0.243 | 0.917 | 0.078 | 0.949 | 0.304 | 0.937 | 0.095 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 1 * model\_\_n\_estimators: 72 * model\_\_max\_depth: 3 * model\_\_min\_samples\_split: 8 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.956 | 0.221 | 0.926 | 0.072 | 0.955 | 0.268 | 0.945 | 0.086 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 33 * model\_\_max\_depth: 10 * model\_\_min\_samples\_split: 5 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.968 | 0.194 | 0.947 | 0.065 | 0.948 | 0.741 | 0.906 | 0.089 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 46 * model\_\_max\_depth: 4 * model\_\_min\_samples\_split: 7 * model\_\_min\_samples\_leaf: 4 * model\_\_max\_features: auto * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.96 | 0.203 | 0.93 | 0.065 | 0.954 | 0.266 | 0.945 | 0.085 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 1 * model\_\_n\_estimators: 60 * model\_\_max\_depth: 3 * model\_\_min\_samples\_split: 17 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: auto * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.961 | 0.21 | 0.932 | 0.068 | 0.957 | 0.272 | 0.943 | 0.086 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 60 * model\_\_max\_depth: 2 * model\_\_min\_samples\_split: 16 * model\_\_min\_samples\_leaf: 4 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.95 | 0.237 | 0.916 | 0.075 | 0.954 | 0.286 | 0.942 | 0.09 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 39 * model\_\_max\_depth: 5 * model\_\_min\_samples\_split: 9 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.961 | 0.203 | 0.936 | 0.066 | 0.955 | 0.271 | 0.943 | 0.088 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 1 * model\_\_n\_estimators: 93 * model\_\_max\_depth: 5 * model\_\_min\_samples\_split: 12 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.964 | 0.197 | 0.937 | 0.065 | 0.957 | 0.25 | 0.942 | 0.082 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 1 * model\_\_n\_estimators: 65 * model\_\_max\_depth: 5 * model\_\_min\_samples\_split: 20 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: auto * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.963 | 0.2 | 0.934 | 0.066 | 0.952 | 0.265 | 0.938 | 0.084 |

## Report of inner training in fold 7 of outer Cross Validation

## Report of training in this outer fold

Best model with respect to selected metric is RandomForestClassifier with the following params:

* undersampling\_majority\_class: False
* model\_\_bootstrap: 0
* model\_\_n\_estimators: 81
* model\_\_max\_depth: 9
* model\_\_min\_samples\_split: 8
* model\_\_min\_samples\_leaf: 3
* model\_\_max\_features: sqrt
* model\_\_class\_weight: balanced

### Comparison of all models trained in this outer fold

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Params** | **Comments** | **roc\_auc on inner fold** | **neg\_log\_loss on inner fold** | **average\_precision on inner fold** | **neg\_brier\_score on inner fold** | **roc\_auc on outer fold** | **neg\_log\_loss on outer fold** | **average\_precision on outer fold** | **neg\_brier\_score on outer fold** |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 10 * model\_\_learning\_rate: 0.057 * model\_\_min\_child\_weight: 7 * model\_\_subsample: 0.883 * model\_\_colsample\_bytree: 0.235 * model\_\_scale\_pos\_weight: 8.783 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.952 | 0.236 | 0.917 | 0.078 | 0.945 | 0.706 | 0.952 | 0.072 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 5 * model\_\_learning\_rate: 0.109 * model\_\_min\_child\_weight: 5 * model\_\_subsample: 0.947 * model\_\_colsample\_bytree: 0.189 * model\_\_scale\_pos\_weight: 0.207 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.948 | 0.233 | 0.911 | 0.075 | 0.947 | 0.732 | 0.946 | 0.085 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 9 * model\_\_learning\_rate: 0.114 * model\_\_min\_child\_weight: 7 * model\_\_subsample: 0.928 * model\_\_colsample\_bytree: 0.561 * model\_\_scale\_pos\_weight: 0.116 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.934 | 0.286 | 0.895 | 0.09 | 0.92 | 0.804 | 0.914 | 0.106 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 9 * model\_\_learning\_rate: 0.088 * model\_\_min\_child\_weight: 4 * model\_\_subsample: 0.846 * model\_\_colsample\_bytree: 0.306 * model\_\_scale\_pos\_weight: 1.236 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.951 | 0.226 | 0.915 | 0.073 | 0.931 | 0.773 | 0.922 | 0.094 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 15 * model\_\_learning\_rate: 0.109 * model\_\_min\_child\_weight: 1 * model\_\_subsample: 0.909 * model\_\_colsample\_bytree: 0.446 * model\_\_scale\_pos\_weight: 0.386 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.939 | 0.254 | 0.901 | 0.084 | 0.935 | 0.349 | 0.933 | 0.097 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_1 * model\_\_max\_depth: 10 * model\_\_learning\_rate: 0.262 * model\_\_min\_child\_weight: 4 * model\_\_subsample: 0.873 * model\_\_colsample\_bytree: 0.138 * model\_\_scale\_pos\_weight: 1.756 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.867 | 0.422 | 0.802 | 0.138 | 0.929 | 0.379 | 0.928 | 0.113 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 13 * model\_\_learning\_rate: 0.109 * model\_\_min\_child\_weight: 10 * model\_\_subsample: 0.814 * model\_\_colsample\_bytree: 0.418 * model\_\_scale\_pos\_weight: 4.06 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.957 | 0.212 | 0.921 | 0.07 | 0.948 | 0.729 | 0.942 | 0.08 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 14 * model\_\_learning\_rate: 0.248 * model\_\_min\_child\_weight: 5 * model\_\_subsample: 0.86 * model\_\_colsample\_bytree: 0.257 * model\_\_scale\_pos\_weight: 8.216 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.95 | 0.227 | 0.914 | 0.075 | 0.943 | 0.745 | 0.946 | 0.084 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 9 * model\_\_learning\_rate: 0.163 * model\_\_min\_child\_weight: 3 * model\_\_subsample: 0.906 * model\_\_colsample\_bytree: 0.56 * model\_\_scale\_pos\_weight: 0.268 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.947 | 0.233 | 0.909 | 0.075 | 0.944 | 0.303 | 0.929 | 0.091 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 10 * model\_\_learning\_rate: 0.086 * model\_\_min\_child\_weight: 9 * model\_\_subsample: 0.879 * model\_\_colsample\_bytree: 0.448 * model\_\_scale\_pos\_weight: 1.626 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.956 | 0.215 | 0.923 | 0.07 | 0.95 | 0.727 | 0.948 | 0.081 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 99 * model\_\_max\_depth: 2 * model\_\_min\_samples\_split: 17 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.946 | 0.248 | 0.909 | 0.078 | 0.937 | 0.749 | 0.934 | 0.09 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 81 * model\_\_max\_depth: 9 * model\_\_min\_samples\_split: 8 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.961 | 0.2 | 0.932 | 0.064 | 0.964 | 0.264 | 0.96 | 0.079 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 22 * model\_\_max\_depth: 5 * model\_\_min\_samples\_split: 5 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.958 | 0.207 | 0.923 | 0.066 | 0.955 | 0.716 | 0.957 | 0.078 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 68 * model\_\_max\_depth: 3 * model\_\_min\_samples\_split: 15 * model\_\_min\_samples\_leaf: 4 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.954 | 0.219 | 0.923 | 0.07 | 0.945 | 0.738 | 0.939 | 0.086 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 53 * model\_\_max\_depth: 4 * model\_\_min\_samples\_split: 7 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.953 | 0.221 | 0.92 | 0.071 | 0.958 | 0.256 | 0.954 | 0.077 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 1 * model\_\_n\_estimators: 67 * model\_\_max\_depth: 7 * model\_\_min\_samples\_split: 12 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.954 | 0.217 | 0.922 | 0.069 | 0.964 | 0.286 | 0.96 | 0.083 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 1 * model\_\_n\_estimators: 25 * model\_\_max\_depth: 5 * model\_\_min\_samples\_split: 6 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.957 | 0.21 | 0.924 | 0.067 | 0.965 | 0.264 | 0.956 | 0.081 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 14 * model\_\_max\_depth: 4 * model\_\_min\_samples\_split: 14 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: auto * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.955 | 0.217 | 0.926 | 0.07 | 0.958 | 0.287 | 0.955 | 0.086 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 79 * model\_\_max\_depth: 7 * model\_\_min\_samples\_split: 19 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.956 | 0.212 | 0.924 | 0.068 | 0.962 | 0.292 | 0.954 | 0.087 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 91 * model\_\_max\_depth: 4 * model\_\_min\_samples\_split: 17 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.954 | 0.22 | 0.921 | 0.07 | 0.958 | 0.269 | 0.95 | 0.082 |

## Report of inner training in fold 9 of outer Cross Validation

## Report of training in this outer fold

Best model with respect to selected metric is XGBClassifier with the following params:

* undersampling\_majority\_class: False
* max\_k\_undersampling: 5
* resample\_\_sampling\_strategy: all
* post\_process\_\_option: option\_3
* model\_\_max\_depth: 7
* model\_\_learning\_rate: 0.177
* model\_\_min\_child\_weight: 6
* model\_\_subsample: 0.971
* model\_\_colsample\_bytree: 0.561
* model\_\_scale\_pos\_weight: 0.156
* model\_\_objective: binary:logistic

### Comparison of all models trained in this outer fold

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Params** | **Comments** | **roc\_auc on inner fold** | **neg\_log\_loss on inner fold** | **average\_precision on inner fold** | **neg\_brier\_score on inner fold** | **roc\_auc on outer fold** | **neg\_log\_loss on outer fold** | **average\_precision on outer fold** | **neg\_brier\_score on outer fold** |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 9 * model\_\_learning\_rate: 0.224 * model\_\_min\_child\_weight: 10 * model\_\_subsample: 0.882 * model\_\_colsample\_bytree: 0.288 * model\_\_scale\_pos\_weight: 0.256 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.938 | 0.263 | 0.901 | 0.086 | 0.93 | 0.316 | 0.919 | 0.093 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_3 * model\_\_max\_depth: 7 * model\_\_learning\_rate: 0.177 * model\_\_min\_child\_weight: 6 * model\_\_subsample: 0.971 * model\_\_colsample\_bytree: 0.561 * model\_\_scale\_pos\_weight: 0.156 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.957 | 0.223 | 0.929 | 0.073 | 0.943 | 0.293 | 0.947 | 0.084 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 7 * model\_\_learning\_rate: 0.161 * model\_\_min\_child\_weight: 9 * model\_\_subsample: 0.905 * model\_\_colsample\_bytree: 0.44 * model\_\_scale\_pos\_weight: 0.677 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.948 | 0.236 | 0.913 | 0.076 | 0.948 | 0.287 | 0.944 | 0.085 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_1 * model\_\_max\_depth: 14 * model\_\_learning\_rate: 0.189 * model\_\_min\_child\_weight: 9 * model\_\_subsample: 0.887 * model\_\_colsample\_bytree: 0.312 * model\_\_scale\_pos\_weight: 0.278 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.925 | 0.306 | 0.885 | 0.1 | 0.893 | 0.401 | 0.887 | 0.122 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_2 * model\_\_max\_depth: 5 * model\_\_learning\_rate: 0.164 * model\_\_min\_child\_weight: 4 * model\_\_subsample: 0.85 * model\_\_colsample\_bytree: 0.389 * model\_\_scale\_pos\_weight: 7.749 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.937 | 0.264 | 0.904 | 0.088 | 0.948 | 0.281 | 0.936 | 0.09 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 10 * model\_\_learning\_rate: 0.289 * model\_\_min\_child\_weight: 4 * model\_\_subsample: 0.907 * model\_\_colsample\_bytree: 0.326 * model\_\_scale\_pos\_weight: 0.294 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.944 | 0.253 | 0.907 | 0.083 | 0.952 | 0.277 | 0.947 | 0.085 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_2 * model\_\_max\_depth: 14 * model\_\_learning\_rate: 0.103 * model\_\_min\_child\_weight: 2 * model\_\_subsample: 0.91 * model\_\_colsample\_bytree: 0.136 * model\_\_scale\_pos\_weight: 4.259 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.942 | 0.266 | 0.904 | 0.087 | 0.933 | 0.318 | 0.937 | 0.099 |
| XGBClassifier | * undersampling\_majority\_class: True * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_1 * model\_\_max\_depth: 6 * model\_\_learning\_rate: 0.078 * model\_\_min\_child\_weight: 2 * model\_\_subsample: 0.801 * model\_\_colsample\_bytree: 0.798 * model\_\_scale\_pos\_weight: 0.609 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 392.0 * average prop of minority class before resampling: 0.385 * average size of training set after resampling: 482.4 * average prop of minority class after resampling: 0.5 | 0.916 | 0.316 | 0.874 | 0.1 | 0.928 | 0.351 | 0.929 | 0.11 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 6 * resample\_\_sampling\_strategy: all * post\_process\_\_option: option\_3 * model\_\_max\_depth: 14 * model\_\_learning\_rate: 0.099 * model\_\_min\_child\_weight: 7 * model\_\_subsample: 0.951 * model\_\_colsample\_bytree: 0.593 * model\_\_scale\_pos\_weight: 0.199 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.954 | 0.223 | 0.924 | 0.07 | 0.948 | 0.271 | 0.949 | 0.079 |
| XGBClassifier | * undersampling\_majority\_class: False * max\_k\_undersampling: 5 * resample\_\_sampling\_strategy: minority * post\_process\_\_option: option\_1 * model\_\_max\_depth: 7 * model\_\_learning\_rate: 0.089 * model\_\_min\_child\_weight: 7 * model\_\_subsample: 0.913 * model\_\_colsample\_bytree: 0.331 * model\_\_scale\_pos\_weight: 0.284 * model\_\_objective: binary:logistic | * option: build model with resampling * number of folds: 5 * average size of training set before resampling: 542.8 * average prop of minority class before resampling: 0.444 * average size of training set after resampling: 603.2 * average prop of minority class after resampling: 0.5 | 0.919 | 0.316 | 0.878 | 0.101 | 0.889 | 0.404 | 0.887 | 0.125 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 75 * model\_\_max\_depth: 6 * model\_\_min\_samples\_split: 7 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.954 | 0.23 | 0.922 | 0.073 | 0.974 | 0.224 | 0.964 | 0.066 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 44 * model\_\_max\_depth: 5 * model\_\_min\_samples\_split: 14 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.951 | 0.227 | 0.919 | 0.072 | 0.968 | 0.233 | 0.96 | 0.068 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 1 * model\_\_n\_estimators: 73 * model\_\_max\_depth: 4 * model\_\_min\_samples\_split: 9 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.951 | 0.234 | 0.919 | 0.073 | 0.975 | 0.241 | 0.966 | 0.074 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 26 * model\_\_max\_depth: 4 * model\_\_min\_samples\_split: 7 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.951 | 0.228 | 0.918 | 0.073 | 0.97 | 0.245 | 0.962 | 0.075 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 1 * model\_\_n\_estimators: 84 * model\_\_max\_depth: 10 * model\_\_min\_samples\_split: 5 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: auto * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.947 | 0.25 | 0.914 | 0.08 | 0.966 | 0.251 | 0.958 | 0.072 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 1 * model\_\_n\_estimators: 62 * model\_\_max\_depth: 4 * model\_\_min\_samples\_split: 12 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.953 | 0.228 | 0.922 | 0.073 | 0.972 | 0.258 | 0.964 | 0.076 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 31 * model\_\_max\_depth: 8 * model\_\_min\_samples\_split: 8 * model\_\_min\_samples\_leaf: 3 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.949 | 0.244 | 0.916 | 0.078 | 0.979 | 0.221 | 0.969 | 0.064 |
| RandomForestClassifier | * undersampling\_majority\_class: False * model\_\_bootstrap: 0 * model\_\_n\_estimators: 58 * model\_\_max\_depth: 7 * model\_\_min\_samples\_split: 7 * model\_\_min\_samples\_leaf: 4 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 542.8 * average prop of minority class: 0.444 | 0.951 | 0.233 | 0.919 | 0.075 | 0.976 | 0.231 | 0.966 | 0.067 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 0 * model\_\_n\_estimators: 46 * model\_\_max\_depth: 5 * model\_\_min\_samples\_split: 6 * model\_\_min\_samples\_leaf: 4 * model\_\_max\_features: sqrt * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.951 | 0.23 | 0.922 | 0.074 | 0.963 | 0.252 | 0.958 | 0.071 |
| RandomForestClassifier | * undersampling\_majority\_class: True * model\_\_bootstrap: 1 * model\_\_n\_estimators: 18 * model\_\_max\_depth: 7 * model\_\_min\_samples\_split: 15 * model\_\_min\_samples\_leaf: 2 * model\_\_max\_features: auto * model\_\_class\_weight: balanced\_subsample | * option: build model without resampling * number of folds: 5 * average size of training set: 392.0 * average prop of minority class: 0.385 | 0.943 | 0.251 | 0.905 | 0.08 | 0.964 | 0.245 | 0.955 | 0.068 |