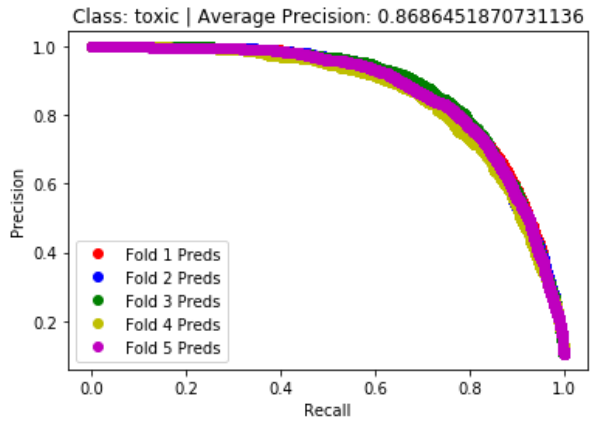
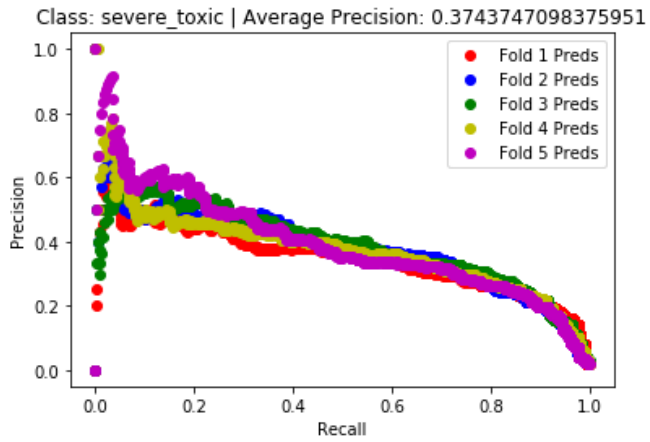
LGB with TF-IDF features (max_features=20000, ngram_range=(1,2) for word, ngram_range=(2,6) for char)

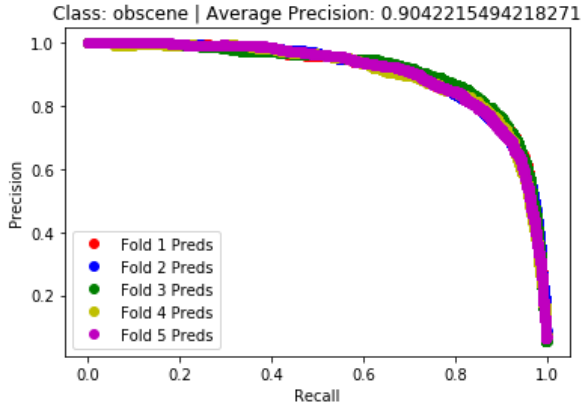
fit: toxic | mean AUC: 0.9715



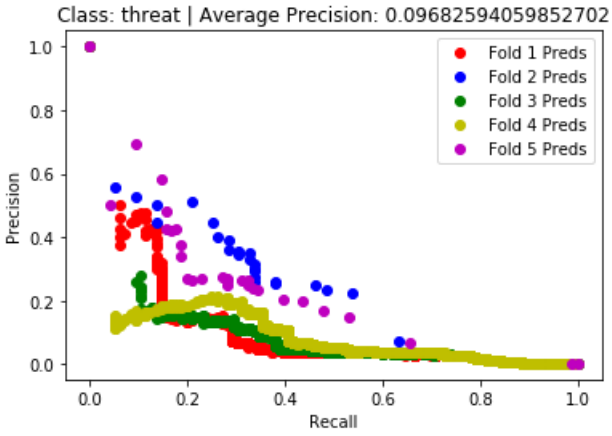
fit: severe_toxic | mean AUC: 0.9814



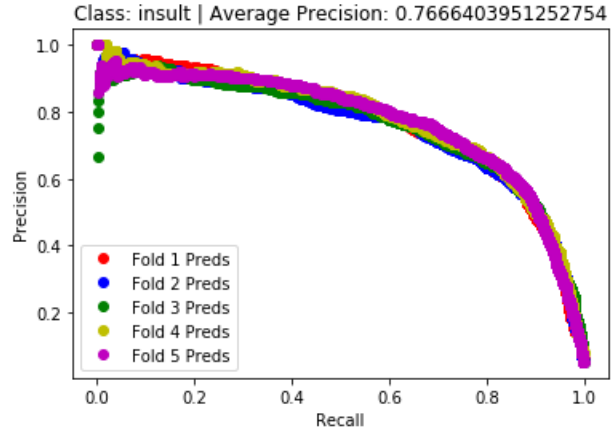
fit: obscene | mean AUC: 0.9904



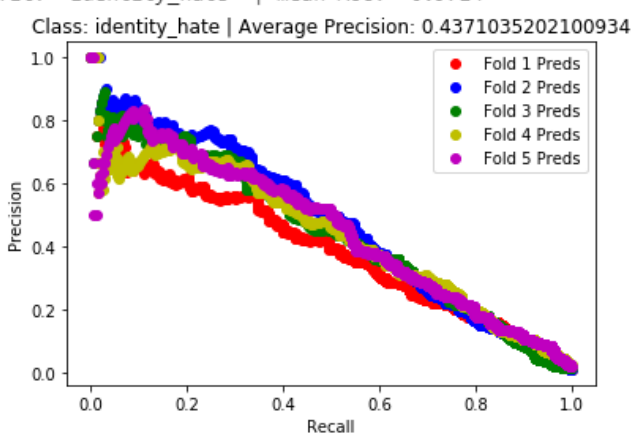
fit: threat | mean AUC: 0.8212



fit: insult | mean AUC: 0.9771

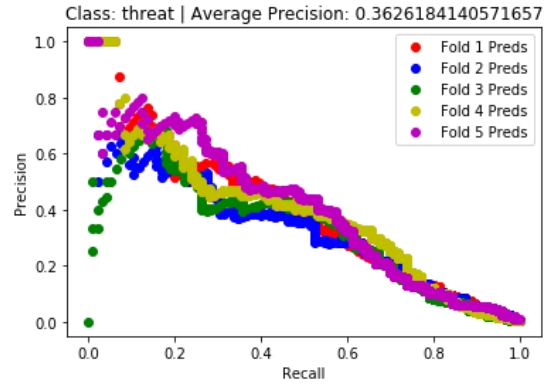


fit: identity_hate | mean AUC: 0.9714

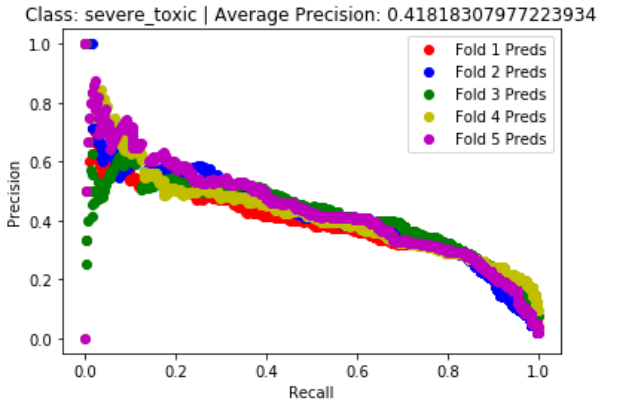


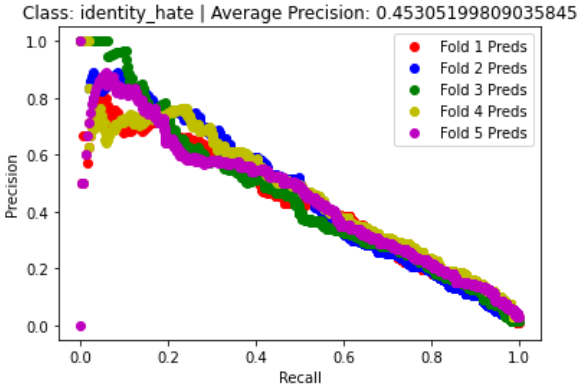
NB-SVM with only word TF-IDF features (ngram_range=(1,2), max_features=20000) & Hyperopt

```
BEST PARAMS:  {'C': 0.9561729886528385, 'tol': 1e-05}
fit:  threat  | mean AUC:  0.9752
```




```
BEST PARAMS:  {'C': 0.30966350550722643, 'tol': 1e-07}
fit:  severe_toxic  | mean AUC:  0.9859
```



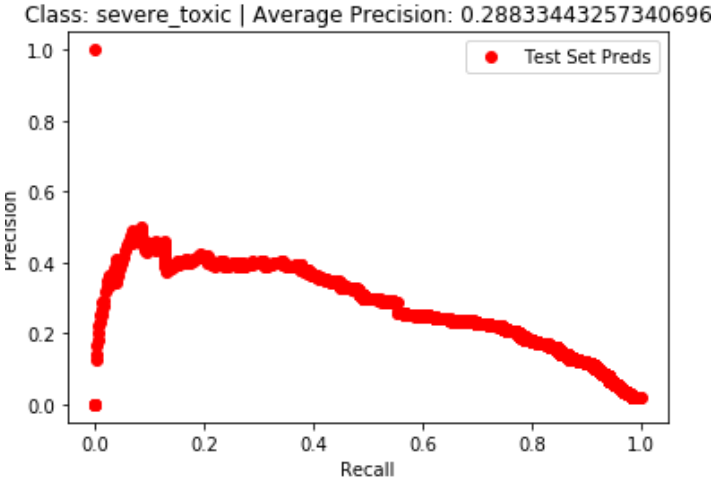


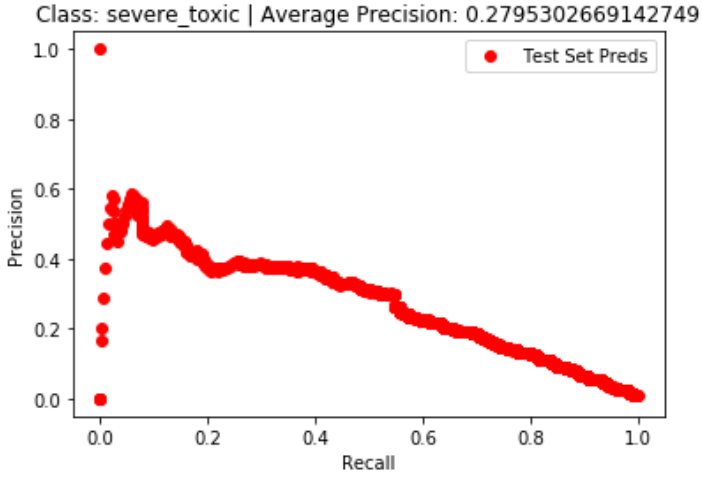
Bi-directional LSTM

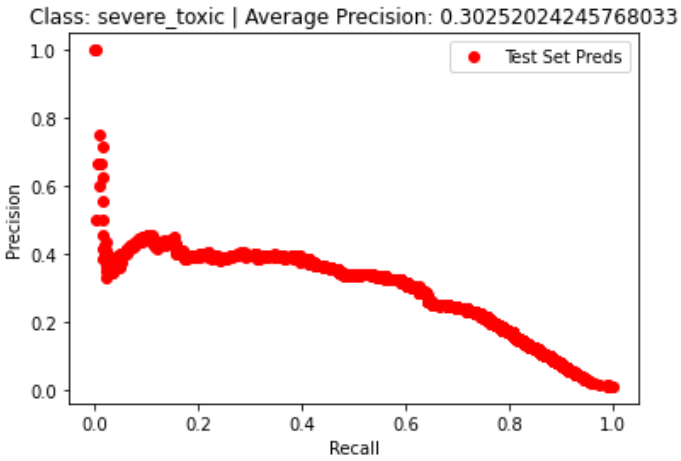
[Tokenizer (20000) -> Embedding (100) -> Bi-directional LSTM (100) -> Dropout (0.2) -> Dense (Sigmoid)

Epochs = 10, batch_size = 64, validation_split = 0.2]

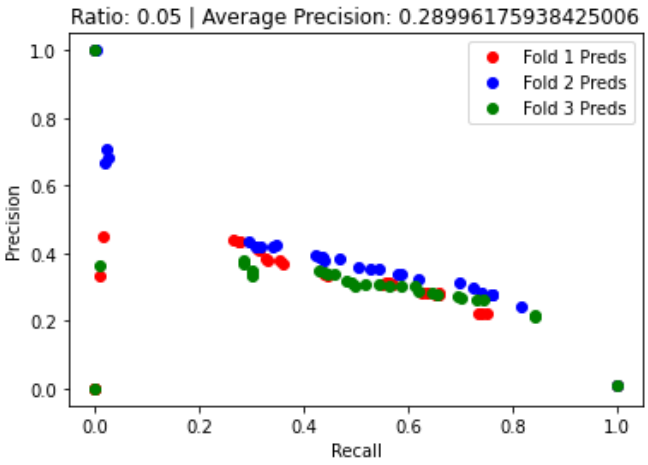
lstm_units = 50

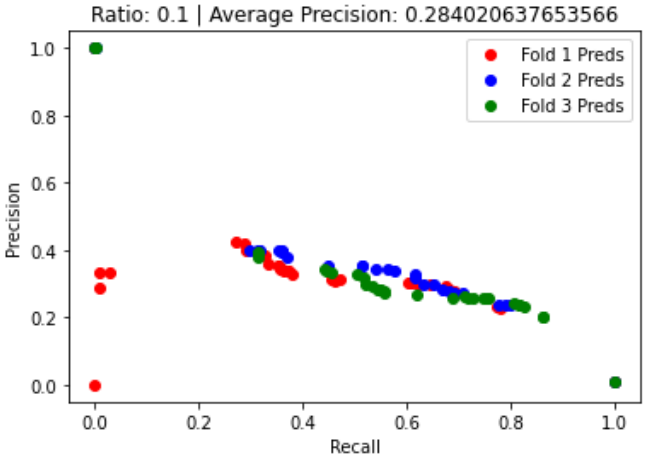




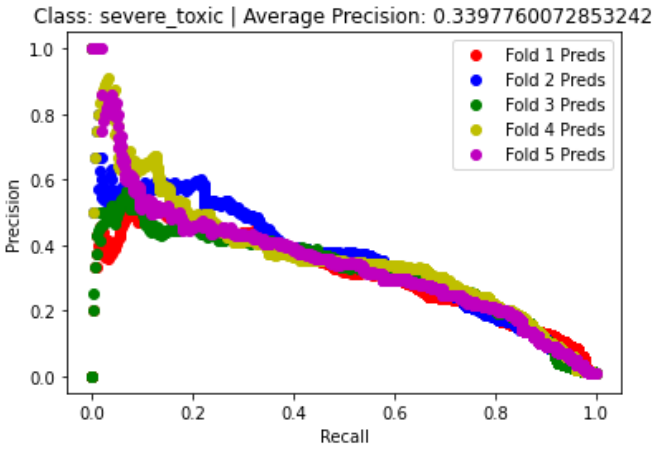


LightGBM With TFIDF Features and SVMSMote and TomekLinks (Severe_Threat)

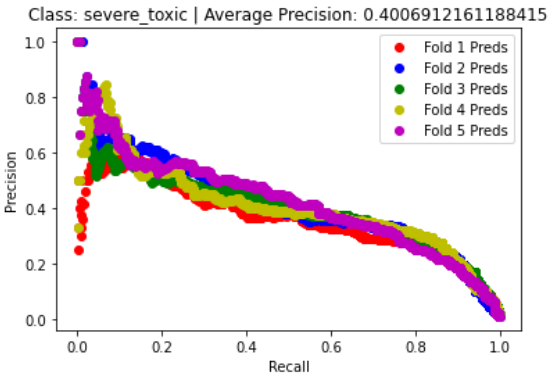




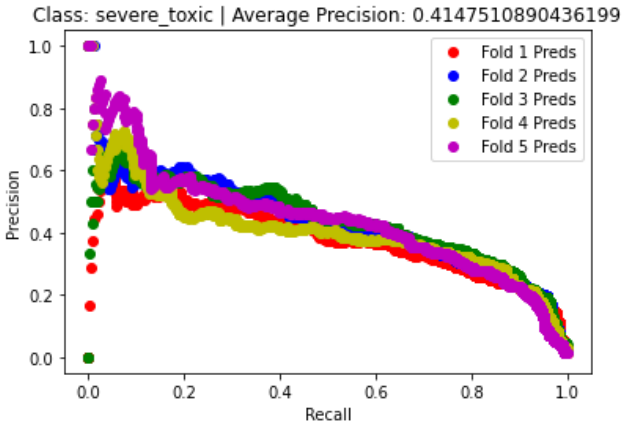
LGB w/ TFIDF + bagging/feature fraction = 0.8



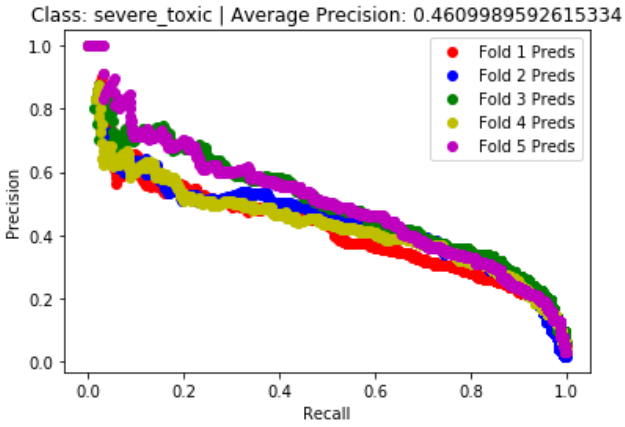
XGB w/ TFIDF + bagging/feature fraction = 0.8 and scale_pos_weight



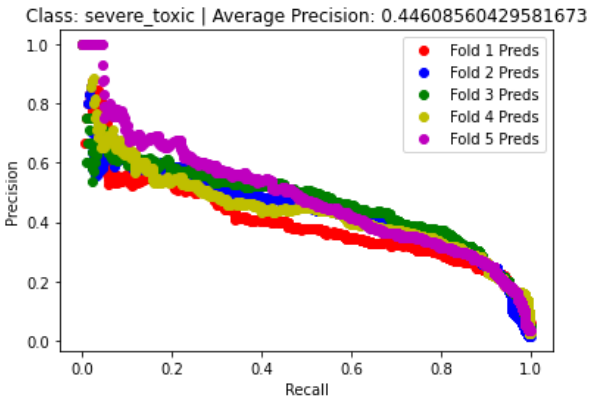
XGB w/ TFIDF + bagging/feature fraction = 0.8 and NO scale_pos_weight,xgb_rounds=500



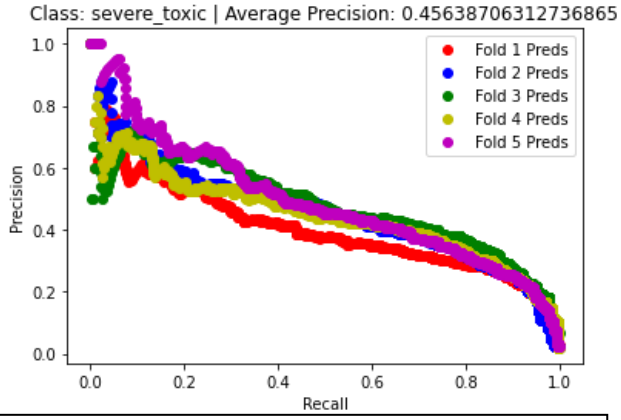
XGB w/ TFIDF + bagging/feature fraction = 0.7 and metadata features (pos + spam)



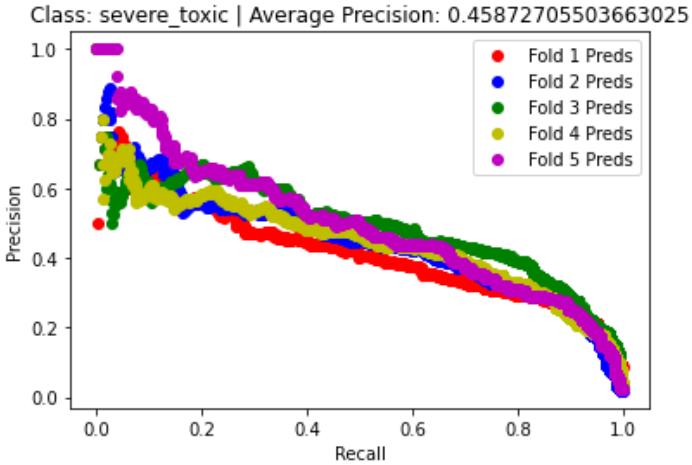
XGB w/ TFIDF + bagging/feature fraction = 0.7 and metadata features (pos + spam) + SMOTE(0.1,kn=5)



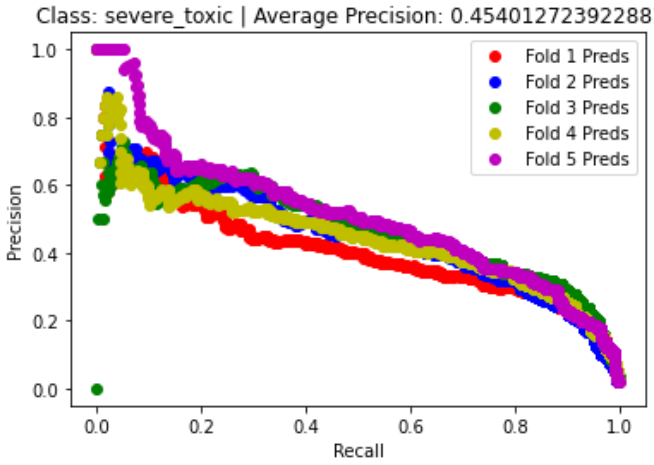
XGB w/ TFIDF + bagging/feature fraction = 0.7 and metadata features (pos + spam) + SMOTE(0.05,kn=5)



XGB w/ TFIDF + bagging/feature fraction = 0.7 and metadata features (pos + spam) + SMOTE(0.05,kn=10)

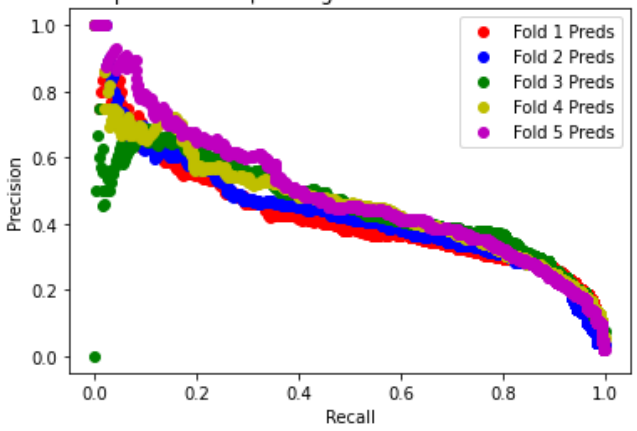


XGB w/ TFIDF + bagging/feature fraction = 0.7 and metadata features (pos + spam) + SMOTE(0.025,kn=10)



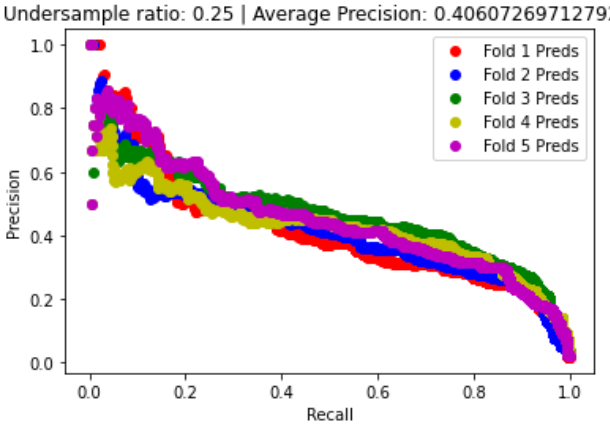
```
Ratio:  0.1  | mean AUC:  0.9872
```

Undersample ratio: 0.1 | Average Precision: 0.43594132463890384



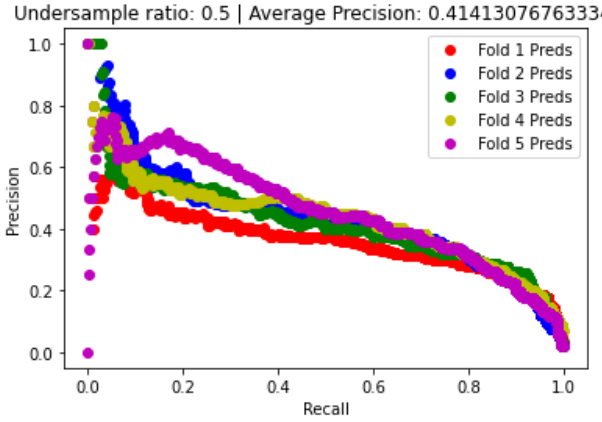

```
Ratio:  0.25  | mean AUC:  0.9854
```

Undersample ratio: 0.25 | Average Precision: 0.40607269712792576



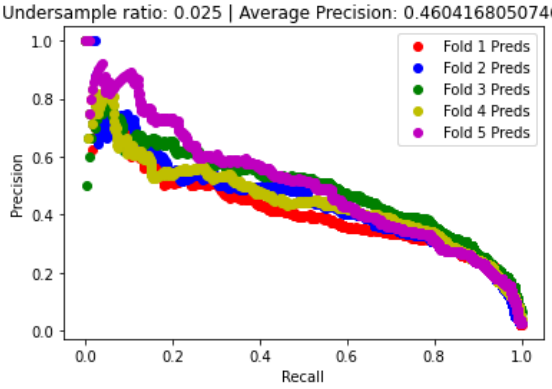

```
Ratio:  0.5  | mean AUC:  0.9856
```

Undersample ratio: 0.5 | Average Precision: 0.4141307676333418




```
Ratio:  0.025  | mean AUC:  0.987
```

Undersample ratio: 0.025 | Average Precision: 0.46041680507464483

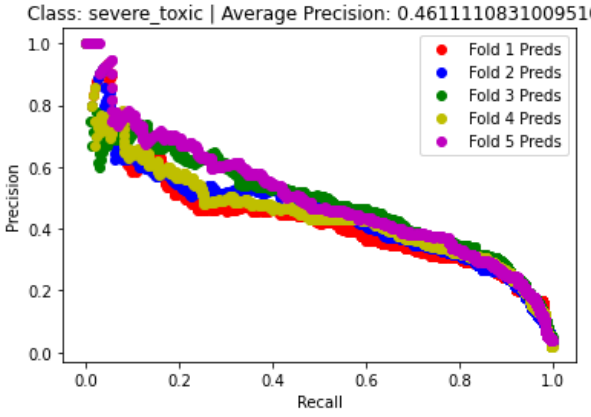


XGB w/ TFIDF + bagging/feature fraction = 0.7 and metadata features (pos + spam + added swearwords)

XGB w/ TFIDF + bagging/feature fraction = 0.7 and metadata features (pos + spam + added swearwords – NO LEMMATIZATION)

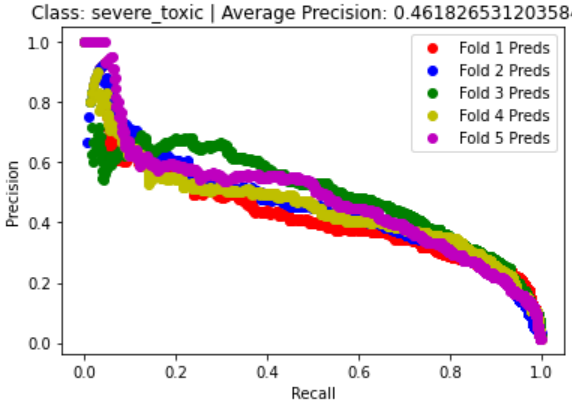
```
Best Threshold=0.251440, F-Score=0.472
Best Threshold=0.200090, F-Score=0.503
Best Threshold=0.207257, F-Score=0.525
Best Threshold=0.190769, F-Score=0.505
Best Threshold=0.186424, F-Score=0.508
Class:  severe_toxic  | mean AUC:  0.9868
Avg best threshold:  0.2072
```

Class: severe_toxic | Average Precision: 0.46111108310095106



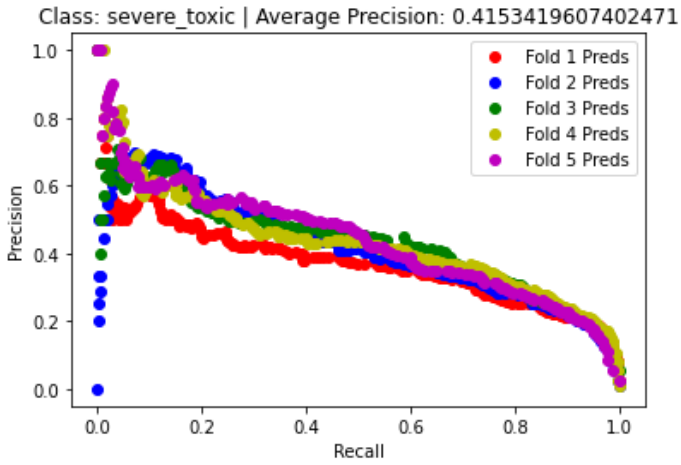

```
Class:  severe_toxic  | mean AUC:  0.9863
Class:  severe_toxic  | mean F-Score:  0.5056
Avg best threshold:  0.1778
```

Class: severe_toxic | Average Precision: 0.4618265312035841



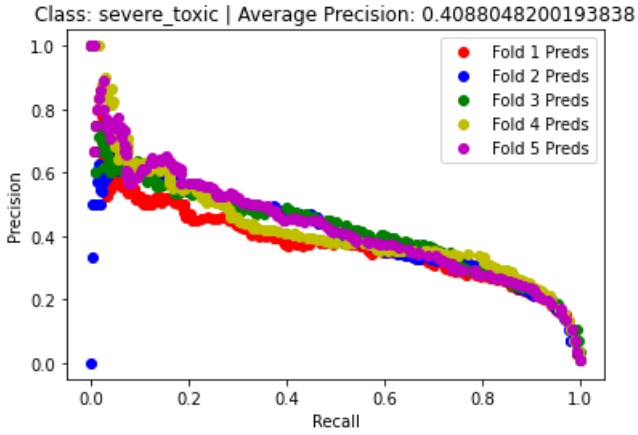
RandomForest w/ TFIDF + num_estimators=200 + SMOTE(0.05,kn=10) and metadata features (pos + spam + added swearwords – NO LEMMATIZATION)

```
Class:  severe_toxic  | mean AUC:  0.9858
Class:  severe_toxic  | mean F-Score:  nan
Avg best threshold:  0.495
```



RandomForest w/ TFIDF + num_estimators=200 + NO SMOTE and metadata features (pos + spam + added swearwords – NO LEMMATIZATION)

```
Class:  severe_toxic  | mean AUC:  0.9838
Class:  severe_toxic  | mean F-Score:  nan
Avg best threshold:  0.343
```



XGB w/ TFIDF + bagging/feature fraction = 0.7 and metadata features (pos + spam + added swearwords – NO LEMMATIZATION)

```
--------- FINAL TEST CLASSIFICATION REPORT ---------
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     63611
           1       0.30      0.17      0.21       367

    accuracy                           0.99     63978
   macro avg       0.65      0.58      0.60     63978
weighted avg       0.99      0.99      0.99     63978

Class:  severe_toxic  | Final AUC:  0.8435
Class:  severe_toxic  | Final F-Score:  0.2133
Best threshold:  0.0513
```