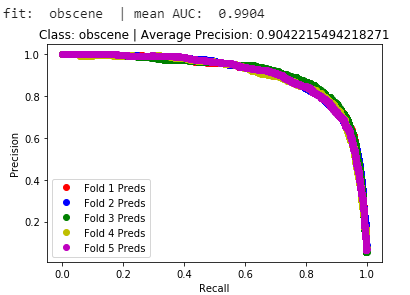
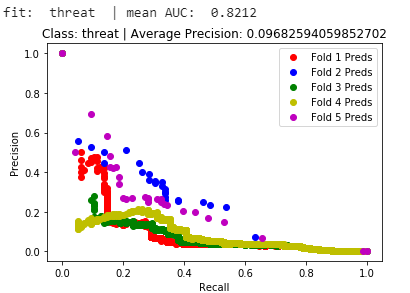
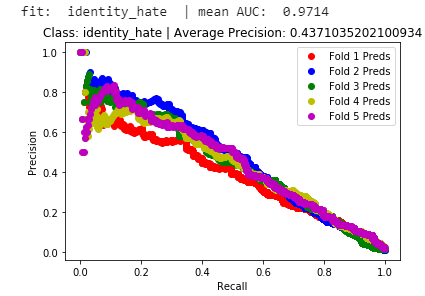
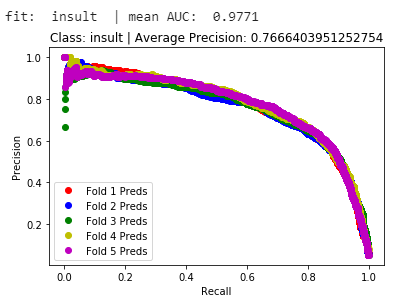
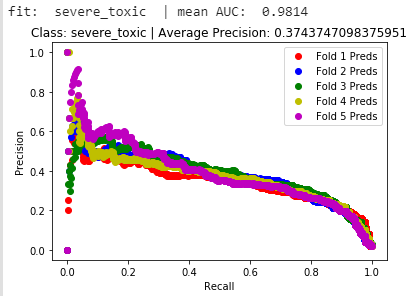
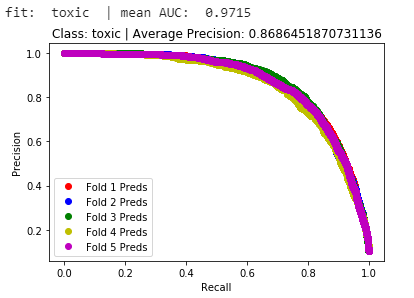
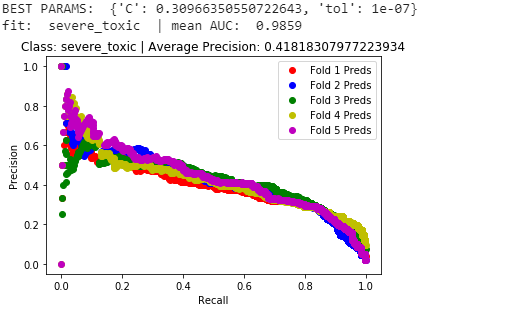
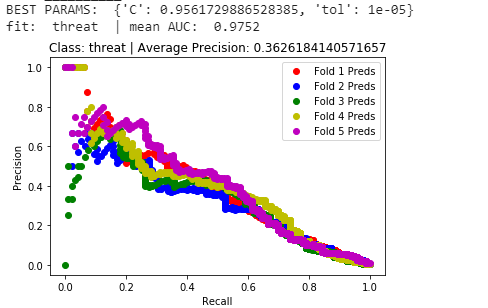
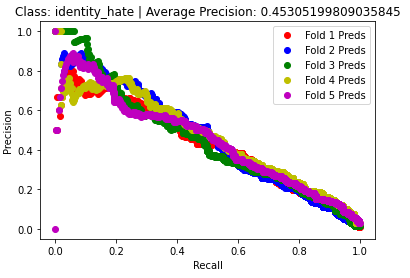
LGB with TF-IDF features (max\_features=20000, ngram\_range=(1,2) for word, ngram\_range=(2,6) for char)

NB-SVM with only word TF-IDF features (ngram\_range=(1,2), max\_features=20000) & Hyperopt



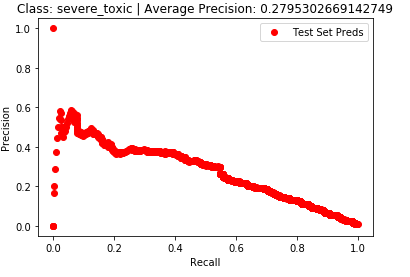
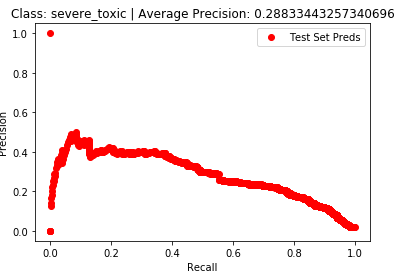


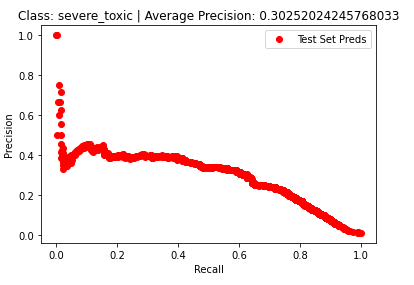
**Bi-directional LSTM**

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

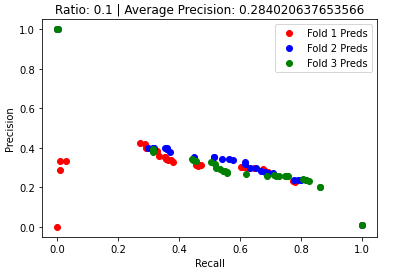
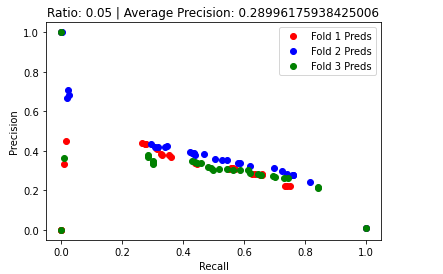
Epochs = 10, batch\_size = 64, validation\_split = 0.2]

lstm\_units = 50





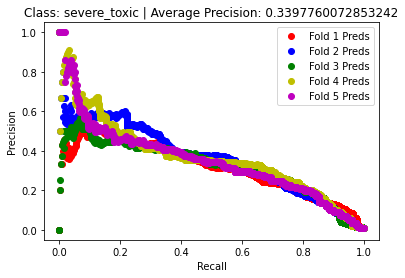
LightGBM With TFIDF Features and SVMSMote and TomekLinks (Severe\_Threat)

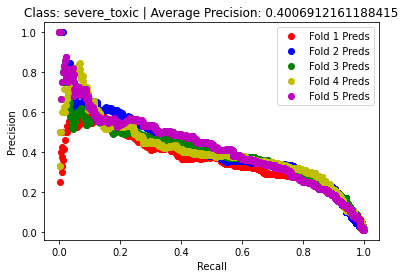


XGB w/ TFIDF + bagging/feature fraction = 0.8

and scale\_pos\_weight

LGB w/ TFIDF + bagging/feature fraction = 0.8



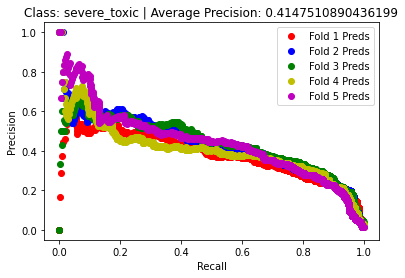
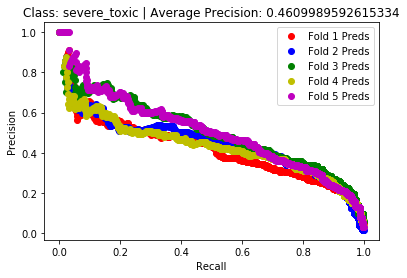


XGB w/ TFIDF + bagging/feature fraction = 0.8

and NO scale\_pos\_weight,xgb\_rounds=500

XGB w/ TFIDF + bagging/feature fraction = 0.7

and metadata features (pos + spam)

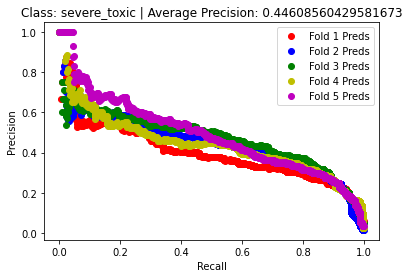
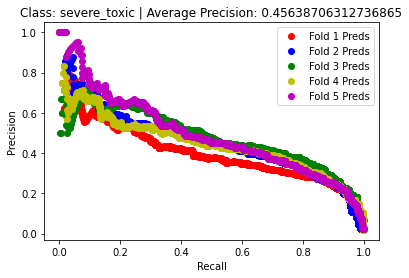


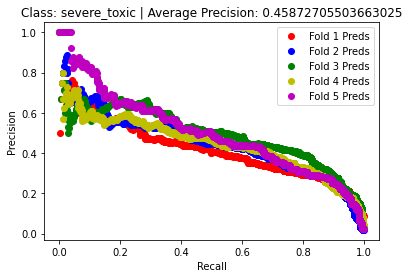
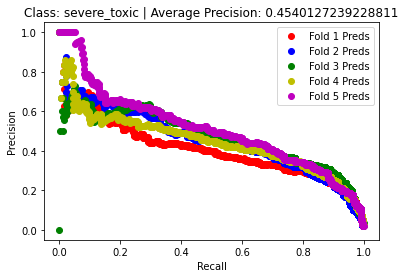
XGB w/ TFIDF + bagging/feature fraction = 0.7

and metadata features (pos + spam) + SMOTE(0.05,kn=5)

XGB w/ TFIDF + bagging/feature fraction = 0.7

and metadata features (pos + spam) + SMOTE(0.1,kn=5)



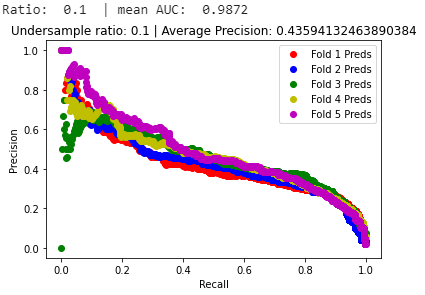
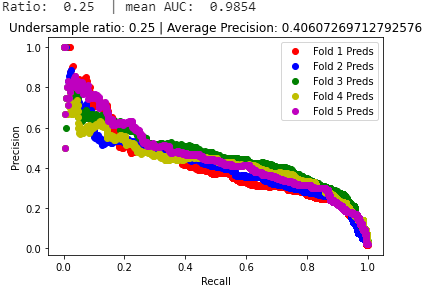


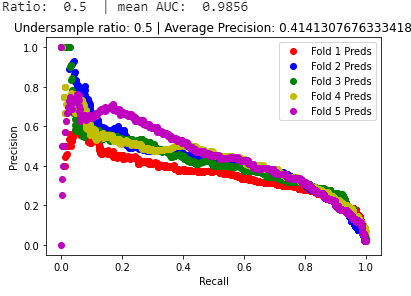
XGB w/ TFIDF + bagging/feature fraction = 0.7

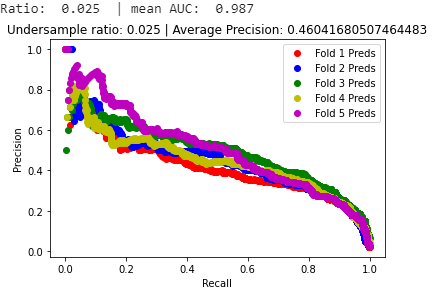
and metadata features (pos + spam) + SMOTE(0.025,kn=10)

XGB w/ TFIDF + bagging/feature fraction = 0.7

and metadata features (pos + spam) + SMOTE(0.05,kn=10)





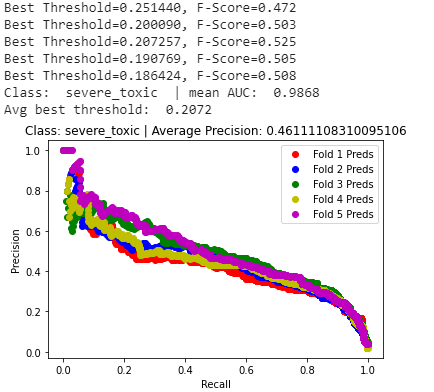


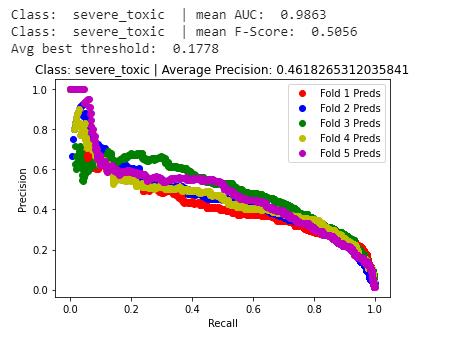
XGB w/ TFIDF + bagging/feature fraction = 0.7

and metadata features (pos + spam + added swearwords)

XGB w/ TFIDF + bagging/feature fraction = 0.7

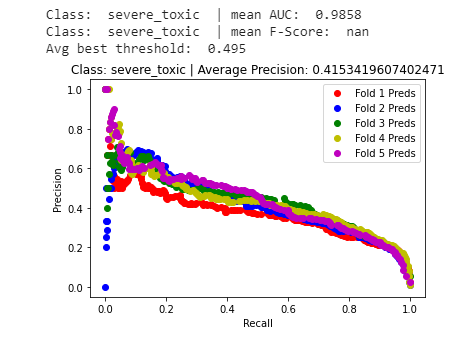
and metadata features (pos + spam + added swearwords – NO LEMMATIZATION)

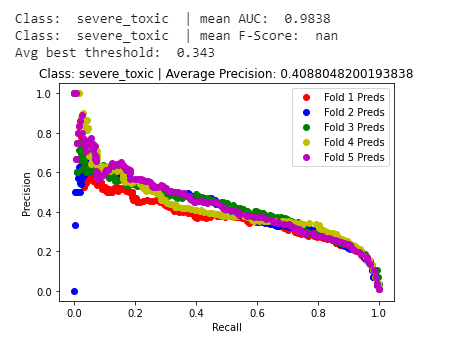
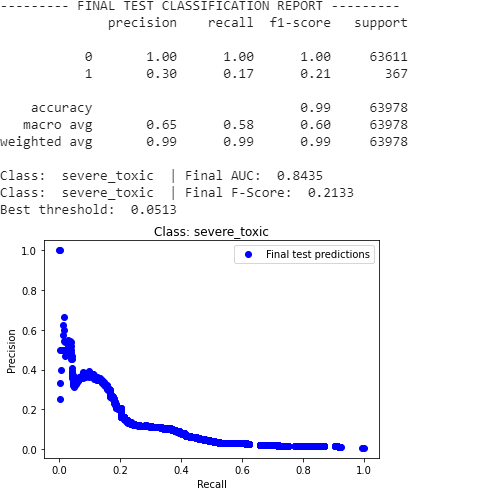




RandomForest w/ TFIDF + num\_estimators=200 + NO SMOTE and metadata features (pos + spam + added swearwords – NO LEMMATIZATION)

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





XGB w/ TFIDF + bagging/feature fraction = 0.7

and metadata features (pos + spam + added swearwords – NO LEMMATIZATION)