Predicting Persuasiveness of Comments

The discussion texts are noisy so first I preprocessed the data and then extract the features.

Preprocess :

Remove Null Text data row
Remove text between sqaure brackets
Apply contractions
Remove stop words
Remove Punctuations
Remove Digit
Apply WordNetLemmatizer(Lemmatization is the process of grouping together the different inflected forms of a word so they can be analysed as a single item. Lemmatization is similar to stemming but it brings context to the words. So it links words with similar meaning to one word.)

Features:

Length
Cosine Similarity between OP and CC
Sentiment score of comment
Number of hedge words in comment

Additional Feature:

User score (Author's score for given comment has already been given in data set)
Number of Urls in comment(Mention of url in comment implies that a citation or evidence has been given for a claim that has been made)

Result:

Confusion Matrix:

```
[[49387   855]
 [ 8704   680]]
              precision    recall  f1-score   support

           0       0.85      0.98      0.91     50242
           1       0.44      0.07      0.12      9384

    accuracy                           0.84     59626
   macro avg       0.65      0.53      0.52     59626
```

weighted avg	0.79	0.84	0.79	59626

Accuracy score = 0.8396840304565123