

AI text Detector using natural language processing

Kalpa Henadhira Arachchige

09.25.2023

GA Project 3





Outline


- ❖ Problem statement
- ❖ Collecting data
- ❖ Preprocessing and EDA
- ❖ Models with CountVectorizer
- ❖ Models with tfidfvectorizer
- ❖ Conclusion

Problem statement

- ❖ To develop an AI text detector for distinguishing between human-generated and AI-generated text with increased accuracy



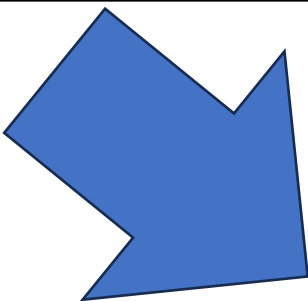
Collecting Data

- 
- science
 - relationship_advice
 - funny
 - NoStupidQuestions
 - AskReddit
 - gaming
 - unresolvedmysteries
 - wewantplates
 - disneyvacation
 - talesfromretail
 - antiMLM
 - IDontWorkHereLady
 - nevertellmetheodds
 - publicfreakout



Subreddits

	question	human_answer	ai_answer
0	Men's shoulder-to-hip ratios influence neuroph...	Y'all, the point isn't that they confirmed tha...	. Men who have shoulder-to-hip ratios of 0.9 o...
1	Pro-circle arguments for a new futuristic city...	Who would willingly live in a super long priso...	1. This new futuristic city will create an unp...
2	Researchers have successfully transferred a ge...	For those unfamiliar, tobacco is a plant that\n\nYes, this is possible. Scientists have us...
3	Boosting the 'warm glow' feeling that people e...	I think there's missing information in that he...	Warm-glow messaging can be used to encourage r...
4	Social myths on nuclear waste being targeted i...	Great news. The war against nuclear power, fun...	1. Nuclear waste is too dangerous to store saf...



Human : 0
AI text : 1

	question	Answer	human_ai
0	Men's shoulder-to-hip ratios influence neuroph...	Y'all, the point isn't that they confirmed tha...	0
1	Pro-circle arguments for a new futuristic city...	Who would willingly live in a super long priso...	0
2	Researchers have successfully transferred a ge...	For those unfamiliar, tobacco is a plant that ...	0
3	Boosting the 'warm glow' feeling that people e...	I think there's missing information in that he...	0
4	Social myths on nuclear waste being targeted i...	Great news. The war against nuclear power, fun...	0
...
12407	I mean he certainly didn't go down without a f...	, but in the end, he was no match for his oppo...	1
12408	In Argentina they capture a thief, tie him up ...	In Argentina, the police will typically appreh...	1
12409	3 guys jump out of a car and rob a lady. (Chic...	This is a very serious crime and would be inve...	1
12410	We're all breaking down...	We're all breaking down in different ways. Som...	1
12411	Karen Go Home	Karen, it's time for you to go home. It's gett...	1

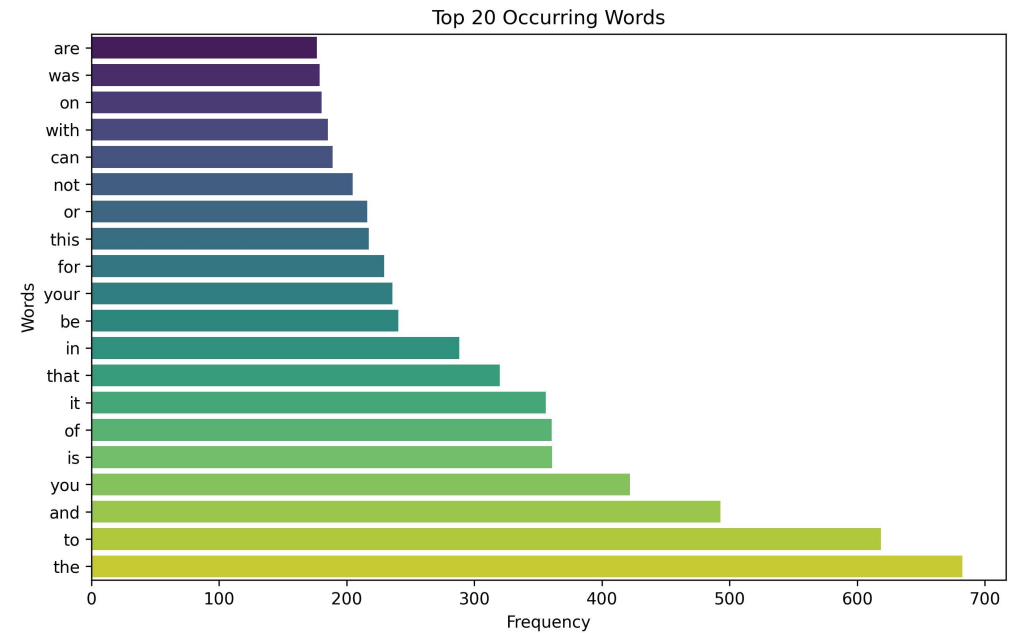
Optimizing Model Features: Key Considerations for Success

- ❖ Common stopwords may be important
- ❖ Identify frequently occurring words
- ❖ Most frequently mentioned terms

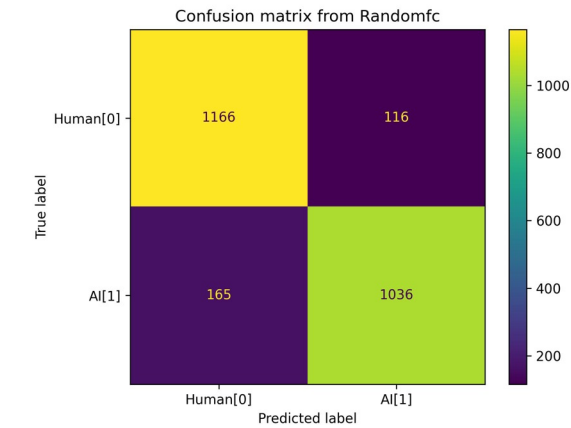
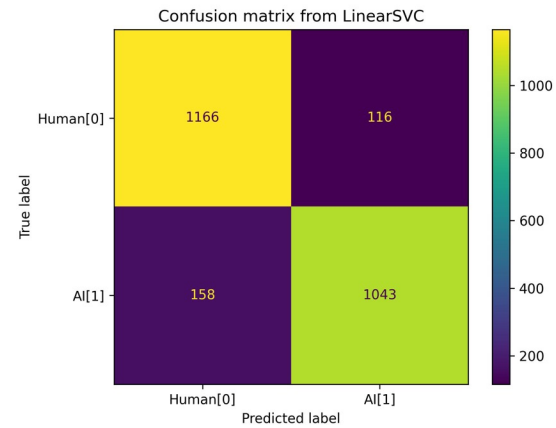
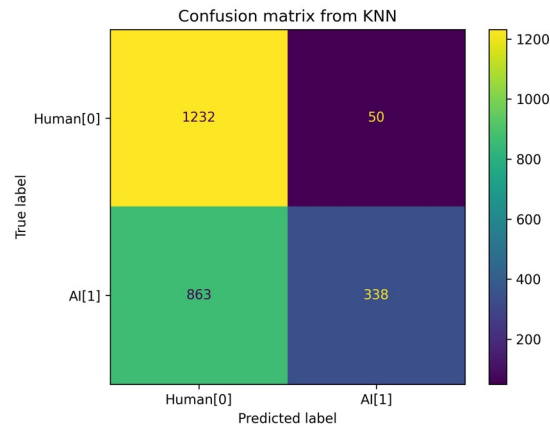
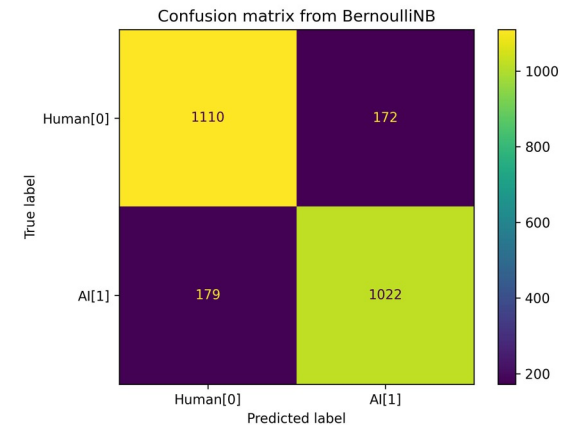
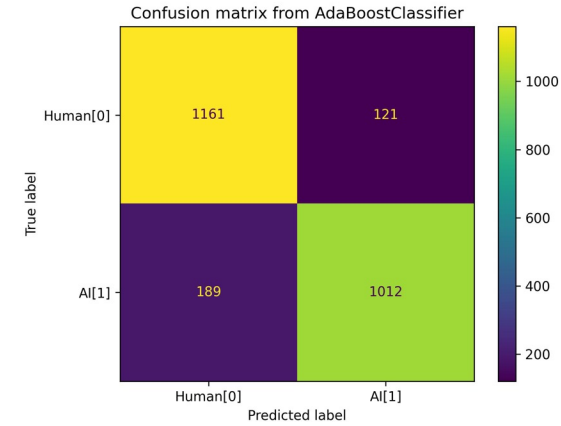
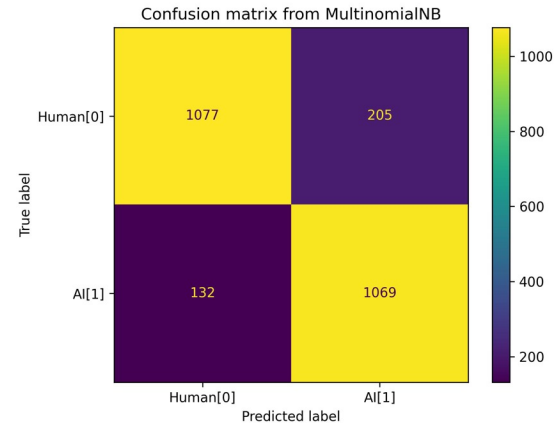
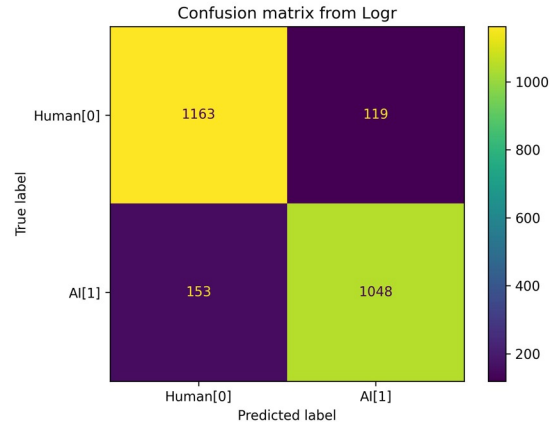
Models with TF-IDFVectorizer

The models that we are going to use here:

- LogisticRegressionCV
- MultinomialNB
- BernoulliNB
- KNN Classifier
- RandomForest Classifier
- AdaBoost Classifier
- SVM Classifier



Confusion matrix with TF-IDFVectorizer



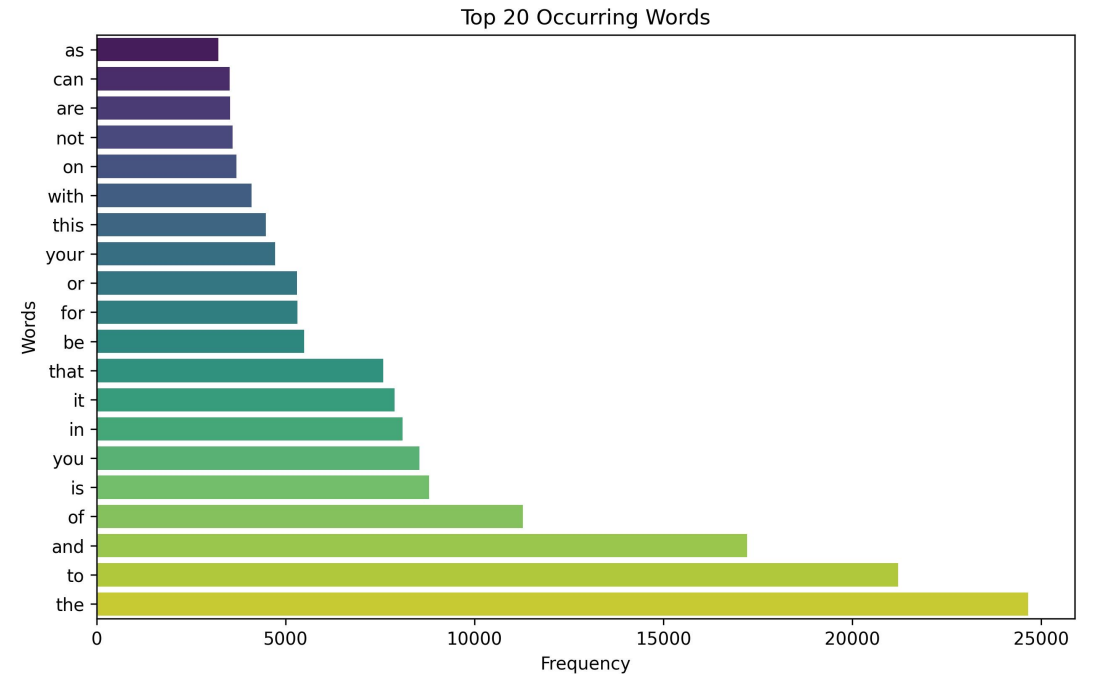
Model Evaluation : TF-IDFVectorizer

	Score on train	Score on test	Sensitivity	Specificity	Precision
Model					
logr	0.951	0.890	0.873	0.907	0.898
Randomfc	0.995	0.887	0.863	0.910	0.899
KNN	0.654	0.632	0.281	0.961	0.871
MultinomialNB	0.893	0.864	0.890	0.840	0.839
BernoulliNB	0.878	0.859	0.851	0.866	0.856
AdaBoostClassifier	0.940	0.875	0.843	0.906	0.893
LinearSVC	0.968	0.890	0.868	0.910	0.900

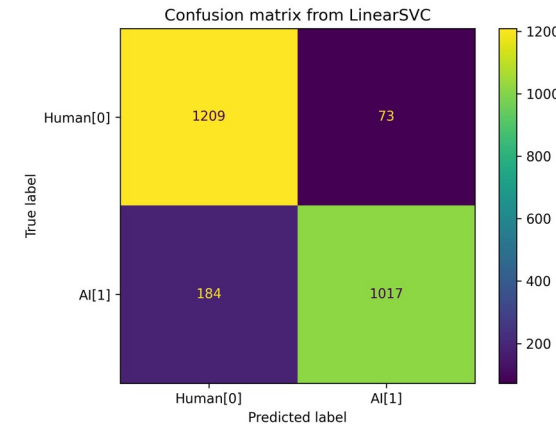
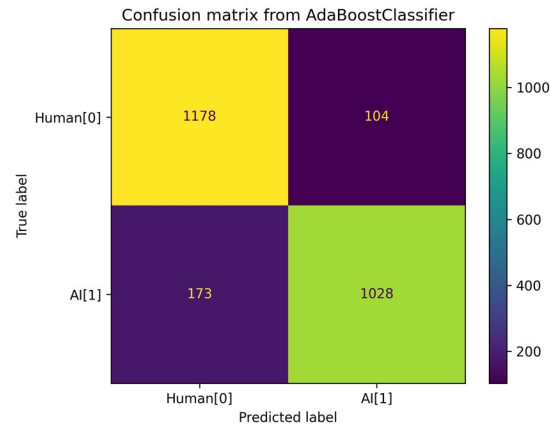
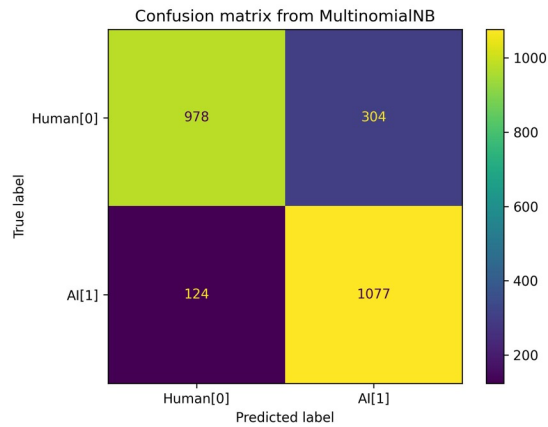
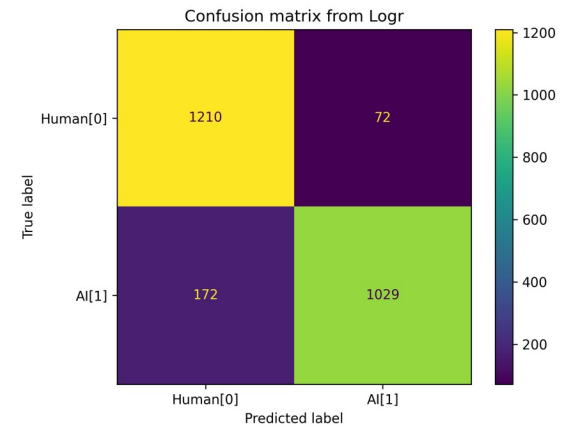
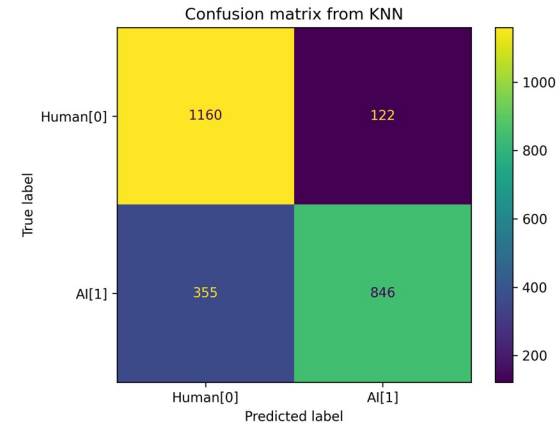
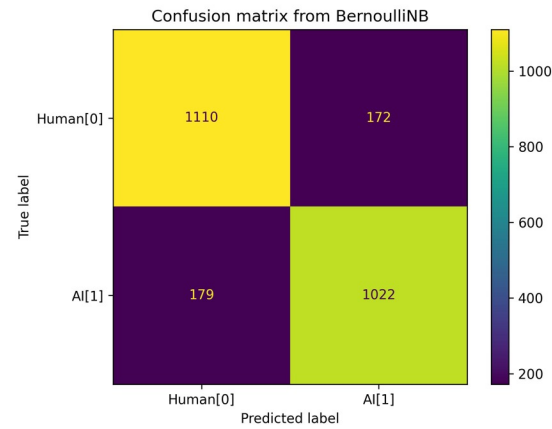
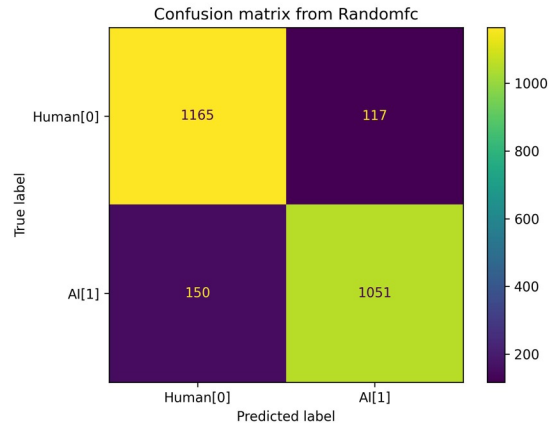
Models with CountVectorizer

The models that we are going to use here:

- LogisticRegressionCV
- MultinomialNB
- BernoulliNB
- KNN Classifier
- RandomForest Classifier
- AdaBoost Classifier
- SVM Classifier



Confusion matrix with CountVectorizer

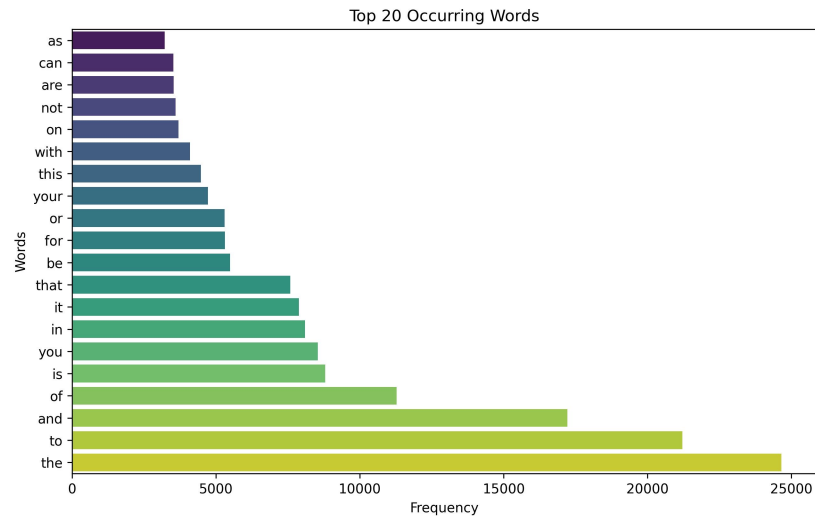


Model Evaluation : CountVectorizer

	Score on train	Score on test	Sensitivity	Specificity	Precision
Model					
logr	0.964	0.902	0.857	0.944	0.935
Randomfc	0.997	0.892	0.875	0.909	0.900
KNN	0.882	0.808	0.704	0.905	0.874
MultinomialNB	0.852	0.828	0.897	0.763	0.780
BernoulliNB	0.878	0.859	0.851	0.866	0.856
AdaBoostClassifier	0.925	0.888	0.856	0.919	0.908
LinearSVC	0.967	0.896	0.847	0.943	0.933

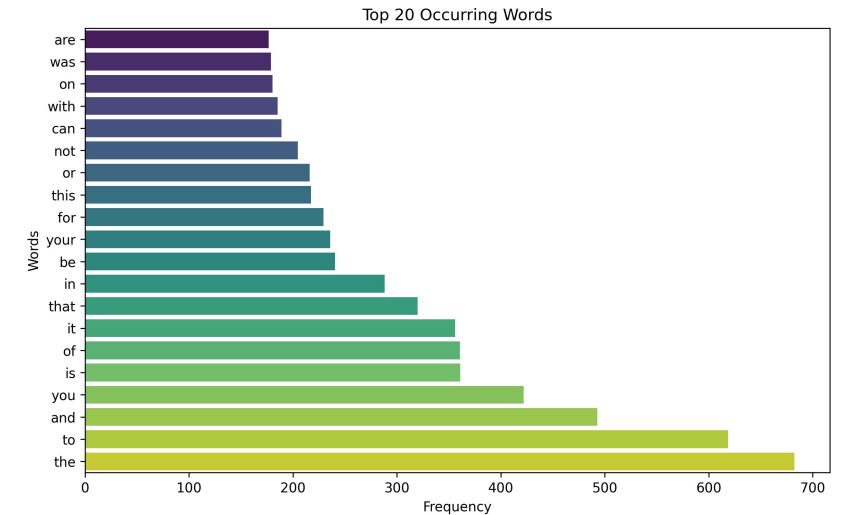
Model comparison

Countvectorizer



	Score on train	Score on test	Sensitivity	Specificity	Precision
Model					
logr	0.964	0.902	0.857	0.944	0.935
Randomfc	0.997	0.892	0.875	0.909	0.900
KNN	0.882	0.808	0.704	0.905	0.874
MultinomialNB	0.852	0.828	0.897	0.763	0.780
BernoulliNB	0.878	0.859	0.851	0.866	0.856
AdaBoostClassifier	0.925	0.888	0.856	0.919	0.908
LinearSVC	0.967	0.896	0.847	0.943	0.933

Tfidfvectorizer



	Score on train	Score on test	Sensitivity	Specificity	Precision
Model					
logr	0.951	0.890	0.873	0.907	0.898
Randomfc	0.995	0.887	0.863	0.910	0.899
KNN	0.654	0.632	0.281	0.961	0.871
MultinomialNB	0.893	0.864	0.890	0.840	0.839
BernoulliNB	0.878	0.859	0.851	0.866	0.856
AdaBoostClassifier	0.940	0.875	0.843	0.906	0.893
LinearSVC	0.968	0.890	0.868	0.910	0.900

Conclusion

- ❖ Our findings indicate that retaining stopwords within our model enhances the accuracy of the AI text detector. Remarkably, the highest accuracy is achieved when employing the logistic regression with cross-validation (LogisticRegressionCV) model.

```
gs_logr.best_params_
```

```
{'tvec__max_df': 0.9,  
  'tvec__max_features': 5000,  
  'tvec__min_df': 3,  
  'tvec__ngram_range': (1, 2),  
  'tvec__stop_words': None}
```

```
gs_logr.best_params_
```

```
{'cvec__max_df': 0.9,  
  'cvec__max_features': 5000,  
  'cvec__min_df': 5,  
  'cvec__ngram_range': (1, 2),  
  'cvec__stop_words': None}
```