

In [1]:

```
!pip install autoviml
```

executed in 7.67s, finished 15:30:22 2022-04-14

Requirement already satisfied: autoviml in c:\users\rani\anaconda3\lib\site-packages (0.1.684)
Requirement already satisfied: nltk in c:\users\rani\anaconda3\lib\site-packages (from autoviml) (3.6.5)
Requirement already satisfied: pandas in c:\users\rani\anaconda3\lib\site-packages (from autoviml) (1.3.4)
Requirement already satisfied: emoji in c:\users\rani\anaconda3\lib\site-packages (from autoviml) (1.7.0)
Requirement already satisfied: ipython in c:\users\rani\anaconda3\lib\site-packages (from autoviml) (7.29.0)
Requirement already satisfied: scikit-learn>=0.23.1 in c:\users\rani\anaconda3\lib\site-packages (from autoviml) (1.0.2)
Requirement already satisfied: beautifulsoup4 in c:\users\rani\anaconda3\lib\site-packages (from autoviml) (4.10.0)
Requirement already satisfied: regex in c:\users\rani\anaconda3\lib\site-packages (from autoviml) (2021.8.3)
Requirement already satisfied: xgboost>=1.1.1 in c:\users\rani\anaconda3\lib\site-packages (from autoviml) (1.5.2)
Requirement already satisfied: matplotlib in c:\users\rani\anaconda3\lib\site-packages (from autoviml) (3.5.2)

In [2]:

```
import pandas as pd
```

executed in 3.26s, finished 15:31:16 2022-04-14

In [5]:

```
data = pd.read_csv('amazonreviews.csv')
data
```

executed in 26ms, finished 15:32:51 2022-04-14

Out[5]:

Unnamed: 0	ratings	reviews
0	0	2
1	1	5
2	2	5
3	3	5
4	4	1
...
1304	1304	4
1305	1305	5
1306	1306	4
1307	1307	5
1308	1308	5

1309 rows × 3 columns

In [6]:

```
data.drop('Unnamed: 0',axis='columns',inplace=True)
```

executed in 6ms, finished 15:34:52 2022-04-14

In [15]:

```
data['target']= data['ratings'].apply(lambda x:1 if x>3 else(-1 if x<3 else 0) )
```

executed in 15ms, finished 15:45:19 2022-04-14

In [22]:

data.head(10)

executed in 12ms, finished 15:48:23 2022-04-14

Out[22]:

ratings			reviews	target
		Ok the DSLR Giant Canon have launched 1 more entry level DSLR. Actually i bought this camera for my friend. Does this camera worth 31K. Lets seePROS1. Its easy to use2. Decent battery backup3. Live mode operation is not lagging.CONNS1. While focusing the DSLR is making too much noise and its the kind of annoying that u will lose interest in taking photos.2. No shortcut or assigned button to change F stop (aperture). A manufacturer who is keeping dedicated button for changing the ISO have to thing about the importance of aperture.3. You can't change the focusing point by pressing the round dial in the system, if u press it some other setup will open up.4. The output image quality is actually not mind blowing.5. White balance- Use it if u want to take risk6. Main issue. If u use camera flash to snap a pic during focusing it will flicker like anything and it will annoy and irritate the subject7. Noise Noise Noise. Whatever low ISO you use noise will b there and when i went to servic...		
0	2			-1
1	5		Good camera for beginners.	1
2	5		Great camera. Perfect product for young budding photographer..... Useful for in-house, nature, functions, and spare time photography. Take high quality pics, videos , portrait, landscape wildlife, nature pics etc	1
3	5		Really impressed with Amazon surprise for delivering within the 2 nd day from ordered Really packing was excellent and the quality of the material is awesome love to control the camera from mobile through Wi-Fi easy to connect, best one for taking self photos and group pics also the quality of canon speaks everything dont go for any oscillation beginners blindly got for it	1
4	1		You are on your own if you buy this product because canon local service center is of no help even in providing the product demo. We logged the complain on canon India site and not bothered to reply. Amazon call raised its hand saying "proof of non service" from local service center. So we have a DSLR that is producing blur images and canon is not providing any proper service. stay away	-1
5	1		Request for replacement of camera main body There is lot of blackspot/blackness automatically appears in View finder mirror just after 15 days.It is requested to please replace the camera as soon as possible.Awaiting early reply.Regards.	-1
6	2		I have received my product DSLR Cannon 1500D at 15 the October 2018 . But Adopter has not been given along with the DSLR which is a very common accessories.They must deliver that as soon as possible.Otheewise I gave to complain about it again & again.	-1
7	3		Auto focus is not at all perfect.Minimum aperture is 4.5, so dont expect much in portrait.While focusing manually there is a play for the lense(18-55 mm) so that there is a chance for getting missing ur focus even though if ur made it correctly.Easy to operate compared to Nikkon models.Fast moving subjects cannot be captured as expected due to its poor auto focus performance. It take about 3 to 6 seconds to focus in internal lighting conditions and it fail to focus several times in dim lights.Plastic quality is poor.I got it for Rs.18599. So that it is a value for money product. Don't go for this model if ur willing to pay about 23k. Better go with Nikkon 5300.	0
8	4		If you want to upgrade your camera from point and shoot to a dslr then it is a good camera for you but if you want to start videography then don't buy it auto focus of video is too much bad. If you want a camera for videography then go foe canon 200D or Nikon D5600	1

ratings		reviews	target
9	5	I am really fully content with the product. It reached me within two days with everything content in it and no doubt in a very good condition. It may be somewhat difficult for the beginners but it will not take a long to learn. If you are choosing between canon and Nikon I will suggest you canon because of its real colour capturing and fine focus of its lenses. You can get more slightly more features in Nikon than canon at the same price but that doesn't mean that will surely better than canon. You can look for contrast between canon and Nikon in the internet. I think it will suggest you the same"canon".	1

In [17]:

```
from sklearn.model_selection import train_test_split  
  
from autoviml.Auto_NLP import Auto_NLP  
  
train, test = train_test_split(data, test_size=0.2)
```

executed in 13ms, finished 15:45:36 2022-04-14

In [20]:

```
input_feature, target = "reviews", "target"

train_x, test_x, final, predicted= Auto_NLP(input_feature, train, test,target,score_ty
```

executed in 39.3s, finished 15:47:51 2022-04-14

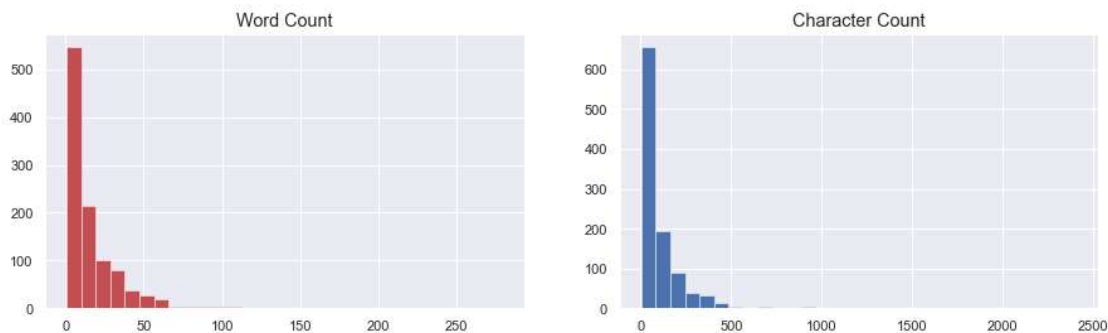
```
[nltk_data] Downloading collection 'popular'
[nltk_data]
[nltk_data] | Downloading package cmudict to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package cmudict is already up-to-date!
[nltk_data] | Downloading package gazetteers to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package gazetteers is already up-to-date!
[nltk_data] | Downloading package genesis to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package genesis is already up-to-date!
[nltk_data] | Downloading package gutenber to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package gutenber is already up-to-date!
[nltk_data] | Downloading package inaugural to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package inaugural is already up-to-date!
[nltk_data] | Downloading package movie_reviews to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package movie_reviews is already up-to-date!
[nltk_data] | Downloading package names to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package names is already up-to-date!
[nltk_data] | Downloading package shakespeare to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package shakespeare is already up-to-date!
[nltk_data] | Downloading package stopwords to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package stopwords is already up-to-date!
[nltk_data] | Downloading package treebank to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package treebank is already up-to-date!
[nltk_data] | Downloading package twitter_samples to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package twitter_samples is already up-to-date!
[nltk_data] | Downloading package omw to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package omw is already up-to-date!
[nltk_data] | Downloading package omw-1.4 to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package omw-1.4 is already up-to-date!
[nltk_data] | Downloading package wordnet to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet is already up-to-date!
[nltk_data] | Downloading package wordnet2021 to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet2021 is already up-to-date!
[nltk_data] | Downloading package wordnet31 to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet31 is already up-to-date!
[nltk_data] | Downloading package wordnet_ic to
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...
```

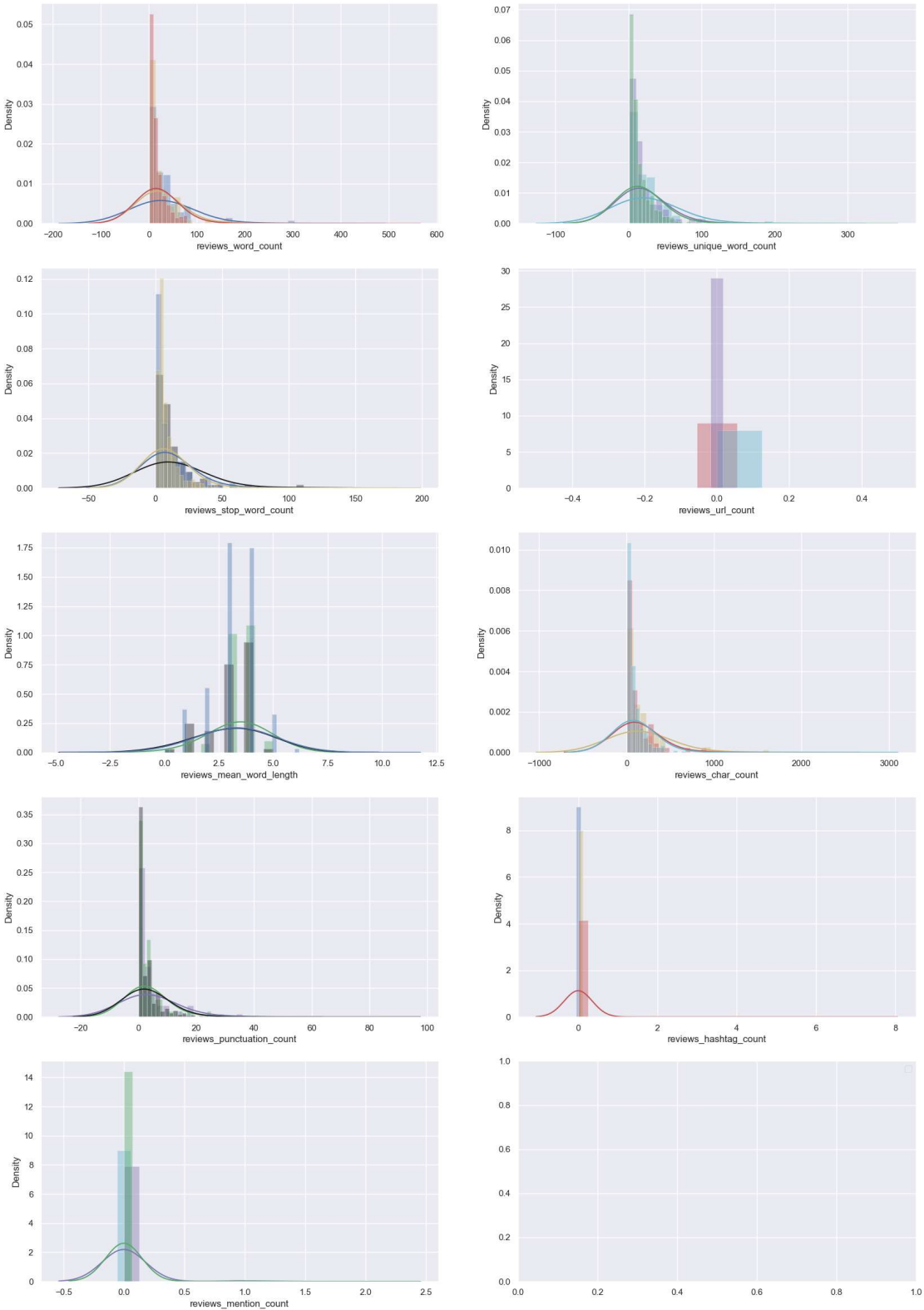
```
[nltk_data] | Package wordnet_ic is already up-to-date!  
[nltk_data] | Downloading package words to  
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...  
[nltk_data] | Package words is already up-to-date!  
[nltk_data] | Downloading package maxent_ne_chunker to  
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...  
[nltk_data] | Package maxent_ne_chunker is already up-to-date!  
[nltk_data] | Downloading package punkt to  
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...  
[nltk_data] | Package punkt is already up-to-date!  
[nltk_data] | Downloading package snowball_data to  
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...  
[nltk_data] | Package snowball_data is already up-to-date!  
[nltk_data] | Downloading package averaged_perceptron_tagger to  
[nltk_data] | C:\Users\Rani\AppData\Roaming\nltk_data...  
[nltk_data] | Package averaged_perceptron_tagger is already up-  
[nltk_data] | to-date!  
[nltk_data] | Done downloading collection popular
```

Auto NLP processing on NLP Column: reviews

Shape of Train Data: 1047 rows

Shape of Test Data: 262 rows





Added 9 summary columns for counts of words and characters in eac

h row

Cleaning text in Train data for reviews column

Faster text processing using clean_text function, since top_num_features is below 300

Time Taken for Expanding emojis and abbreviations in data = 0 seconds

Time Taken for Processing text in data = 5 seconds

Train data Text cleaning completed. Time taken = 13 seconds

For target = target

For class = -1

Top 200 n-grams

```
: ['not', 'good', 'product', 'card', 'memory', 'camera', 'quality',
'but', 'bad', 'amazon', 'buy', 'miss', 'use', 'no', 'picture', 'do',
'bag', 'canon', 'receive', 'purchase', 'me', 'video', 'battery', 'like',
'work', 'box', 'which', 'problem', 'flash', 'lens', 'even', 'ok',
'return', 'defective', 'beginner', 'please', 'go', 'think', 'charge',
'seal', 'when', 'there', 'deliver', 'price', 'they', 'help', 'cannot',
'photo', 'time', 'give', 'dslr', 'image', 'show', 'if', 'light',
'order', 'money', 'blur', 'carry', 'poor', 'properly', 'service', 'also',
'after', 'useful', 'open', 'stop', 'disappointed', 'day', 'low',
'class', 'highly', 'attach', 'your', 'phone', 'replacement', 'waste',
'what', 'as', 'item', 'well', 'include', 'seller', 'now', 'case', 'high',
'complain', 'never', 'want', 'package', 'possible', 'soon', 'happy',
'sd', 'center', 'costly', 'faulty', 'support', 'full', 'month',
'take', 'come', 'don', 'we', 'eos', 'cannon', 'recording', 'could',
'already', 'mobile', 'great', 'mean', 'outing_face', 'p', 'background',
'customer', 'automatically', 'thing', 'focus', 'call', 'send', 'why',
'care', 'today', 'hot', 'will', 'nikon', 'external', 'mistake',
'first', 'much', 'model', 'guy', 'memory card', 'not good', 'do not',
'bad product', 'quality not', 'picture quality', 'camera bag', 'not buy',
'buy canon', 'good but', 'help me', 'soon possible', 'flash light',
'product amazon', 'not work', 'card box', 'buy camera', 'no memory',
'good beginner', 'sd card', 'canon eos', 'not use', 'camera but',
'bag miss', 'also not', 'not receive', 'not charge', 'seal break', 'good quality',
'quality but', 'time but', 'p outing_face', 'open product',
'do not buy', 'camera bag miss', 'as soon possible', 'do not work',
'good quality but', 'memory card box', 'no memory card', 'order product amazon',
'picture quality not', 'quality not good', 'use if want', 'do not do',
'not buy camera', 'please do not', 'buy canon eos', 'card no carry',
'memory card no', 'no carry bag']
```

Top n-grams that are most frequent in this class are: 184

For class = 0

Top 200 n-grams

```
: ['good', 'not', 'camera', 'beginner', 'product', 'but', 'do', 'error',
'intext', 'quality', 'card', 'canon', 'best', 'video', 'focus', 'price',
'picture', 'memory', 'lens', 'buy', 'sd', 'use', 'k', 'after', 'time',
'ok', 'performance', 'take', 'day', 'nice', 'much', 'pic', 'work',
'amazon', 'will', 'shoot', 'auto', 'receive', 'as', 'which', 'no',
'if', 'battery', 'go', 'kit', 'there', 'continuous', 'bag', 'compare',
'also', 'fi', 'wi', 'okay', 'me', 'low', 'photography', 'night',
'want', 'like', 'money', 'hang', 'light', 'external', 'overall',
'photo', 'lense', 'case', 'pay', 'mic', 'normal', 'expect', 'less',
'bad', 'open', 'range', 'body', 'average', 'need', 'tell', 'flash',
'when', 'please', 'review', 'miss', 'image', 'mention', 'shop', 'however',
'support', 'always', 'well', 'help', 'include', 'though', 'basic',
'return', 'automatically', 'box', 'item', 'purpose', 'package', 'mm',
'model', 'experience', 'point', 'mark', 'setting', 'poor', 'give',
'service', 'backup', 'description', 'mobile', 'move', 'dslr', 'ma
```

```
ke', 'even', 'value', 'because', 'cannot', 'second', 'several', 'thin
g', 'portrait', 'internal', 'play', 'we', 'youtube', 'refund', 'perfe
ct', 'back', 'bcoz', 'totally', 'do not', 'good beginner', 'not goo
d', 'memory card', 'kit lens', 'sd card', 'product price', 'auto focu
s', 'but not', 'good camera', 'not buy', 'camera beginner', 'camera b
ody', 'best price', 'nice product', 'after day', 'wi fi', 'picture qu
ality', 'good but', 'not work', 'pic quality', 'value money', 'good p
roduct', 'focus take', 'card camera', 'flash which', 'not receive',
'overall good', 'external mic', 'beginner but', 'good performance',
'several time', 'battery backup', 'do not buy', 'auto focus take', 'q
uality not good', 'memory card camera', 'photography but not', 'do no
t receive', 'auto focus not', 'product do not', 'but do not', 'memory
card bag', 'not receive memory', 'receive memory card']
```

Top n-grams that are most frequent in this class are: 177

For class = 1

Top 200 n-grams

```
: ['good', 'beginner', 'camera', 'product', 'best', 'nice', 'qualit
y', 'picture', 'dslr', 'awesome', 'great', 'price', 'use', 'but', 'li
ke', 'not', 'easy', 'excellent', 'buy', 'focus', 'auto', 'love', 'ama
zon', 'photography', 'canon', 'perfect', 'also', 'really', 'lens', 'i
f', 'range', 'go', 'super', 'budget', 'money', 'purchase', 'learn',
'video', 'photo', 'value', 'well', 'superb', 'happy', 'first', 'k',
'amazing', 'much', 'level', 'low', 'time', 'worth', 'do', 'entry', 'w
ill', 'battery', 'delivery', 'ok', 'handle', 'no', 'work', 'take', 'm
ust', 'image', 'mm', 'mode', 'thanks', 'me', 'start', 'pic', 'card',
'amaze', 'bag', 'need', 'experience', 'option', 'give', 'day', 'wan
t', 'friendly', 'which', 'useful', 'cam', 'light', 'feature', 'everyt
hing', 'could', 'look', 'make', 'autofocus', 'wifi', 'recommend', 'ev
en', 'user', 'high', 'overall', 'red_heart', 'professional', 'photogr
apher', 'deliver', 'as', 'would', 'deal', 'shoot', 'offer', 'prime',
'click', 'satisfy', 'expect', 'mobile', 'carry', 'without', 'basic',
'thank', 'lot', 'who', 'capture', 'memory', 'know', 'receive', 'new',
'because', 'ok_hand', 'night', 'every', 'eyes', 'smiling_face_with_he
art', 'should', 'lense', 'choice', 'point', 'performance', 'little',
'function', 'good beginner', 'picture quality', 'good product', 'came
ra beginner', 'good camera', 'best camera', 'auto focus', 'nice camer
a', 'nice product', 'quality good', 'value money', 'good quality', 'b
est beginner', 'easy use', 'entry level', 'dslr beginner', 'camera go
od', 'best dslr', 'product beginner', 'do not', 'perfect beginner',
'awesome product', 'also good', 'good picture', 'product good', 'pric
e range', 'awesome camera', 'level dslr', 'must buy', 'great produc
t', 'really good', 'first dslr', 'great beginner', 'good camera begin
ner', 'picture quality good', 'best camera beginner', 'entry level ds
lr', 'good picture quality', 'best dslr beginner', 'picture quality a
wesome', 'beginner picture quality', 'good product beginner', 'best e
ntry level', 'picture quality great', 'beginner must buy', 'but do no
t', 'amazing picture quality', 'nice camera beginner', 'nice picture
quality', 'quality also good', 'good entry level', 'beginner like m
e', 'auto focus good', 'good auto focus', 'good camera but', 'best ca
mera range', 'camera price range', 'excellent camera beginner', 'nice
product beginner', 'great picture quality', 'good product but', 'begi
nner learn photography', 'good quality product', 'worth every penny',
'easy use good', 'entry level camera']
```

Top n-grams that are most frequent in this class are: 199

Time Taken = 0 seconds

```
##### THIS IS FOR BUILD_MODEL = TRUE #####
#####
```

Building Model and Pipeline for NLP column = reviews. This will take time...

Class -> Counts -> Percent

```
-1:      97 ->    9.3%
0:      70 ->    6.7%
1:     880 ->   84.0%
```

Selected the maximum number of features limit = 1943

Performing RandomizedSearchCV across 30 params. Optimizing for accuracy

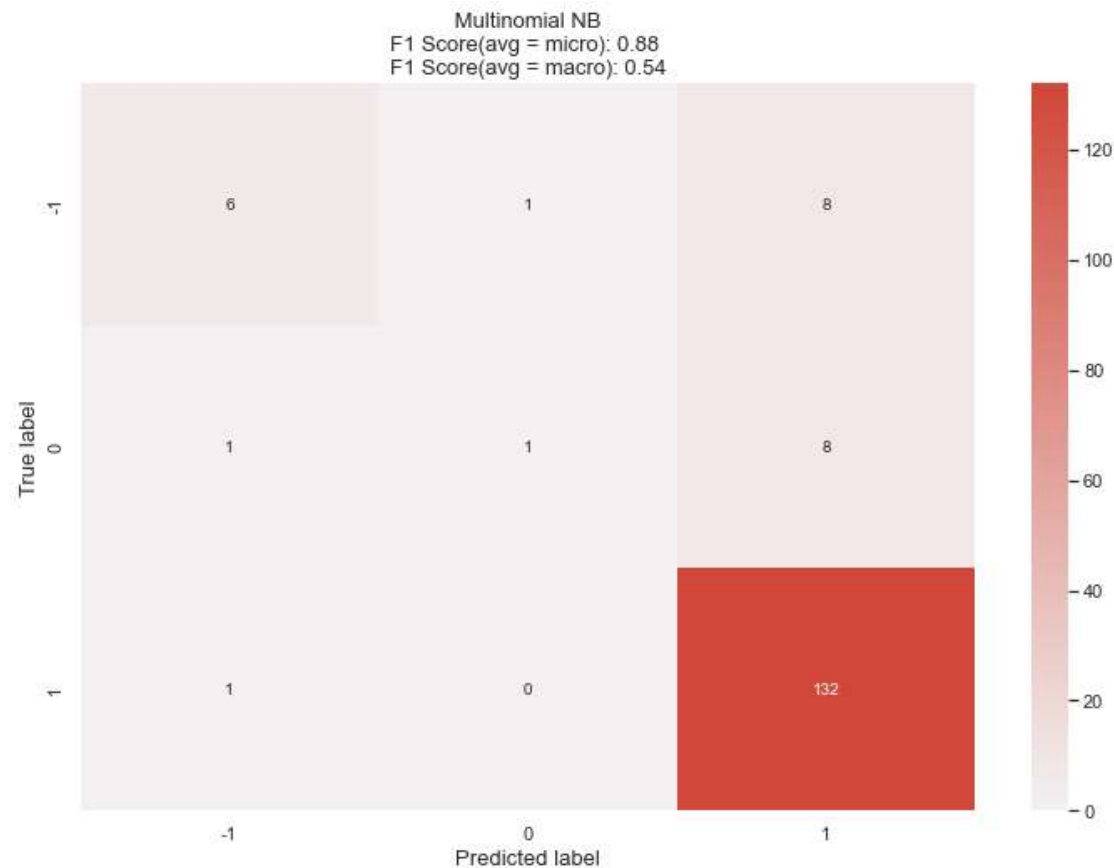
Using train data = (889,) and Cross Validation data = (158,)

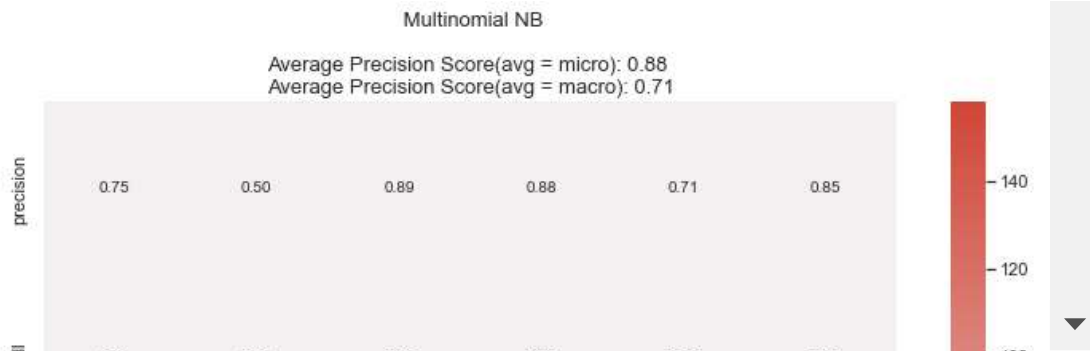
Using a Calibrated Classifier in this Multi-Classification dataset to improve results...

Since top_num_features = 100, Multinomial NB model selected. If you need different model, increase it >= 300.

Training completed. Time taken for training = 0.2 minutes

Best Params of NLP pipeline are: {'multinomialnb__alpha': 0.03266133482097422, 'selectkbest__k': 1438, 'tfidfvectorizer__binary': True, 'tfidfvectorizer__encoding': 'latin-1', 'tfidfvectorizer__max_df': 0.546268635370521}





Now transforming Train data to return as output...

Time Taken for Expanding emojis and abbreviations in data = 0 second

S

Time Taken for Processing text in data = 5 seconds

Transforming Test data to return as output...

Time Taken for Expanding emojis and abbreviations in data = 0 second

S

Time Taken for Processing text in data = 1 seconds

Training best Auto_NLP Pipeline on full Train data...will be faster since best params are known

Time Taken for Expanding emojis and abbreviations in data = 0 second

S

Time Taken for Processing text in data = 4 seconds

Returning best Auto_NLP pipeline to transform and make predictions on test data...

Time Taken for Expanding emojis and abbreviations in data = 0 second

S

Time Taken for Processing text in data = 1 seconds

Training completed. Time taken for Auto_NLP = 0.7 minutes

A U T O N L P C O M P L E T E D #####
#####

In [21]:

```
final.predict(test_x[input_feature])
```

executed in 1.32s, finished 15:47:53 2022-04-14

Time Taken for Expanding emojis and abbreviations in data = 0 second

S

Time Taken for Processing text in data = 1 seconds

Out[21]:

```
array([ 1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1, -1,  1,  1,  1,
 1,
        1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,
 1,
        1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,
 1,
       -1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,
 1,
        1,  1,  1,  1,  1,  1,  1, -1,  1,  1,  1,  1,  1,  1,  1,  1,
 1,
        1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,
 1,
        1,  1,  1, -1,  1,  1,  1,  1,  1,  1, -1, -1, -1,  1,  1,  1, -
 1,
        1,  1,  1,  1,  1,  1,  1, -1, -1,  1,  1,  1,  1,  1,  1,  1,
 1,
        1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1, -1,
 1,
        1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1, -1,  1,  1,  1,  1,
 1,
        1, -1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1, -1,  1,  1, -1,
 1,
       -1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1, -1,  1,  1,  1,
 1,
        1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,
 1,
        1,  1, -1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,
 1,
        1,  1,  1,  1, -1,  1,  1,  1,  1,  1, -1,  1,  1,  1,  1,  1,
 1,
        1, -1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,  1,
 1], dtype=int64)
```

In []: