

BayesAlgorithm

October 21, 2024

```
[2]: import pandas as pd, numpy as np
import matplotlib.pyplot as plt, seaborn as sns
import plotly.express as px
```

```
[3]: import re # regular expression
import nltk # natural language toolkit
import string
```

```
[4]: from nltk.corpus import stopwords
nltk.download('punkt')
nltk.download('stopwords')
```

```
[nltk_data] Downloading package punkt to
[nltk_data] /home/eec795a1-aada-4506-a501-
[nltk_data] 7eee41ea48ce/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package stopwords to
[nltk_data] /home/eec795a1-aada-4506-a501-
[nltk_data] 7eee41ea48ce/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

```
[4]: True
```

```
[5]: from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from wordcloud import WordCloud
stop_words = stopwords.words()
```

0.0.1 Data Inspection

```
[7]: data = pd.read_csv("IMDB Dataset.csv")
data.head()
```

```
[7]:
```

	review	sentiment
0	One of the other reviewers has mentioned that ...	positive
1	A wonderful little production. The...	positive
2	I thought this was a wonderful way to spend ti...	positive

```

3 Basically there's a family where a little boy ... negative
4 Petter Mattei's "Love in the Time of Money" is... positive

```

```
[8]: data.shape
```

```
[8]: (50000, 2)
```

```
[9]: data.isnull().sum()
```

```
[9]: review      0
      sentiment  0
      dtype: int64
```

```
[10]: data.isna().sum()
```

```
[10]: review      0
      sentiment  0
      dtype: int64
```

```
[11]: data.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   review      50000 non-null   object
 1   sentiment   50000 non-null   object
dtypes: object(2)
memory usage: 781.4+ KB

```

```
[12]: data.describe()
```

```

[12]:
count      review sentiment
unique      49582         2
top    Loved today's show!!! It was a variety and not...  positive
freq              5      25000

```

0.0.2 EDA

```
[14]: data['review'].value_counts()
```

```
[14]: Loved today's show!!! It was a variety and not solely cooking (which would have
      been great too). Very stimulating and captivating, always keeping the viewer
      peeking around the corner to see what was coming up next. She is as down to
      earth and as personable as you get, like one of us which made the show all the
```

more enjoyable. Special guests, who are friends as well made for a nice surprise too. Loved the 'first' theme and that the audience was invited to play along too. I must admit I was shocked to see her come in under her time limits on a few things, but she did it and by golly I'll be writing those recipes down. Saving time in the kitchen means more time with family. Those who haven't tuned in yet, find out what channel and the time, I assure you that you won't be disappointed.

5

Hilarious, clean, light-hearted, and quote-worthy. What else can you ask for in a film? This is my all-time, number one favorite movie. Ever since I was a little girl, I've dreamed of owning a blue van with flames and an observation bubble.

The cliché characters in ridiculous situations are what make this film such great fun. The wonderful comedic chemistry between Stephen Furst (Harold) and Andy Tennant (Melio) make up most of my favorite parts of the movie. And who didn't love the hopeless awkwardness of Flynn? Don't forget the airport antics of Leon's cronies, dressed up as Hari Krishnas: dancing, chanting and playing the tambourine--unbeatable! The clues are genius, the locations are classic, and the plot is timeless.

A word to the wise, if you didn't watch this film when you were little, it probably won't win a place in your heart today. But nevertheless give it a chance, you may find that "It doesn't matter what you say, it doesn't matter what you do, you've gotta play."

4

How has this piece of crap stayed on TV this long? It's terrible. It makes me want to shoot someone. It's so fake that it is actually worse than a 1940s sci-fi movie. I'd rather have a stroke than watch this nonsense. I remember watching it when it first came out. I thought, hey this could be interesting, then I found out how absolutely, insanely, ridiculously stupid it really was. It was so bad that I actually took out my pocket knife and stuck my hand to the table.

Please people, stop watching this and all other reality shows, they're the trash that is jamming the networks and canceling quality programming that requires some thought to create.

3

Nickelodeon has gone down the toilet. They have kids saying things like "Oh my God!" and "We're screwed"

This show promotes hate for people who aren't good looking, or aren't in the in crowd. It say that sexual promiscuity is alright, by having girls slobbering over shirtless boys. Not to mention the overweight boy who takes off his shirt. The main characters basically shun anyone out of the ordinary. Carly's friend Sam, who may be a lesbian, beats the snot out of anybody that crosses her path, which says it's alright to be a b**ch. This show has so much negativity in it that nobody should watch it! I give it a 0 out of 10!!!

3

You do realize that you've been watching the EXACT SAME SHOW for eight years, right? I could understand the initial curiosity of seeing strangers co-exist on an Island, but you'd think that after watching unkempt, stink-laden heroes run roughshod through the bush with an egg on a spoon for half a decade would be enough to get you to commit to something a little more original (and

interesting).

And I'm not even speaking of the shows validity which for the record I find questionable. It's just hard to suspend disbelief for "Bushy Bill" eating a rat when the entire crew of producers and camera people are housed in an air conditioned make-shift bio-dome sipping frosty mochcinno's with moxy.

What's the appeal here? I don't care about these people or their meandering lives. I just don't get it. But if you DO find yourself being captivated by hairy, unwashed people, I suggest you turn off your TV and just take a trip to your local bus station where you can see people like this in their TRUE habitat. They call them HOMELESS PEOPLE, and free of charge, you can sit back and marvel in their uncanny ability to retrieve various cigarette debris from a plethora of garbage canisters, eventually striking "pay-dirt" and fashioning a homemade Dr. Frankenstein-styled cancer-stick, all the while begging people for change for food when the stink of "Aqua Velva" on their breath is enough to suggest otherwise. And the best part? Much like Survivor, every week one member of the tribe "Leaves" the "Island" when they are unceremoniously sent packing to the local Institution when the frightening unmedicated state of full-blown schizophrenia kicks into gear! Now THAT'S ENTERTAINMENT! 3

..
REVOLT OF THE ZOMBIES (2 outta 5 stars) No, this is not a long-lost ancestor to the classic George A. Romero zombie flicks. This is a low-budget potboiler from 1936 that probably seemed very cool to audiences of the time... but seems awfully routine these days. There is actually a pretty good scene at the start of a soldier firing off his pistol into a horde of approaching zombie soldiers... and a close-up of bullets entering the bare chest of one of them. The effect looks hopelessly fake these days but in 1936 I'm sure it had audiences gasping. The story concerns the search for the secret of mind control... ostensibly to create an unstoppable zombie army... but later as a means for one character to win the woman he loves. The movie is barely an hour long but moves at a snail's pace so it seems feature-length, believe me! There really isn't much to recommend it... you may get some amusement from the faked studio shots of the star "wading" through a "swamp". The ending is interesting... so I'd say the movie is worth seeing at least once. More than likely you will see it as an extra feature on some cheap "4 movies on 1 DVD" compilation at Wal-Mart for five bucks. Hey, it's well worth the money...

1

At least among those movies with 100 votes or more. Nominated for best screenplay written directly for the screen? Brenda Blethyn nominated for best actress in a leading role?? Nominated for best picture?? I always disagree with many of the Oscar picks, but this movie might very well be the worst movie of all time to be honored by the Academy. The writing and acting were both horrible. Blethyn's performance in particular was one of the worst I've ever seen, and probably the most over-rated acting performance of all time. Awful movie, not worthy of the big screen and not worthy of any cable or television channel that has ever played it, including HBO(where I saw it). I am only thankful I didn't actually pay to see one of the most over-rated movies of all time.

1
 I rented this movie under the impression that it was "Scarecrow 3:Dark Harvest", thinking it was a continuation in the Scarecrow Slayer series (another extremely laughable and all together awful series of movies). I wasn't disappointed though. It was just as awful, if not worse, than what I expected. I was laughing throughout the entire movie. Every piece of bad acting, poorly shot and cut footage, and terrible special effects is what makes this movie worth renting.

The special features include a pathetic view into the cast and crew's six months of filming.

Favorite line, "The sins of my forefathers! They've trickled down to this very moment of time!"

1
 Surreal film noir released soon after the "real," genre-defining classics "The Maltese Falcon," "Double Indemnity" and "The Postman Always Rings Twice." Welles films shouldn't be evaluated against others. He was playing by different rules. In fact, he was playing. This starts where other femme fatale films leave off, so the vaguely logical (but interesting) whodunit is embellished with a display of Wellesian scenes (typical rapid-fire style), dialog (lots of "hard-boiled" philosophy), and unusual acting (good Hayworth presumably intentionally one-dimensional). To Welles "genre" may have meant "formula" but he seemed to like using "mysteries" as backgrounds for his "entertainments."

1
 No one expects the Star Trek movies to be high art, but the fans do expect a movie that is as good as some of the best episodes. Unfortunately, this movie had a muddled, implausible plot that just left me cringing - this is by far the worst of the nine (so far) movies. Even the chance to watch the well known characters interact in another movie can't save this movie - including the goofy scenes with Kirk, Spock and McCoy at Yosemite.

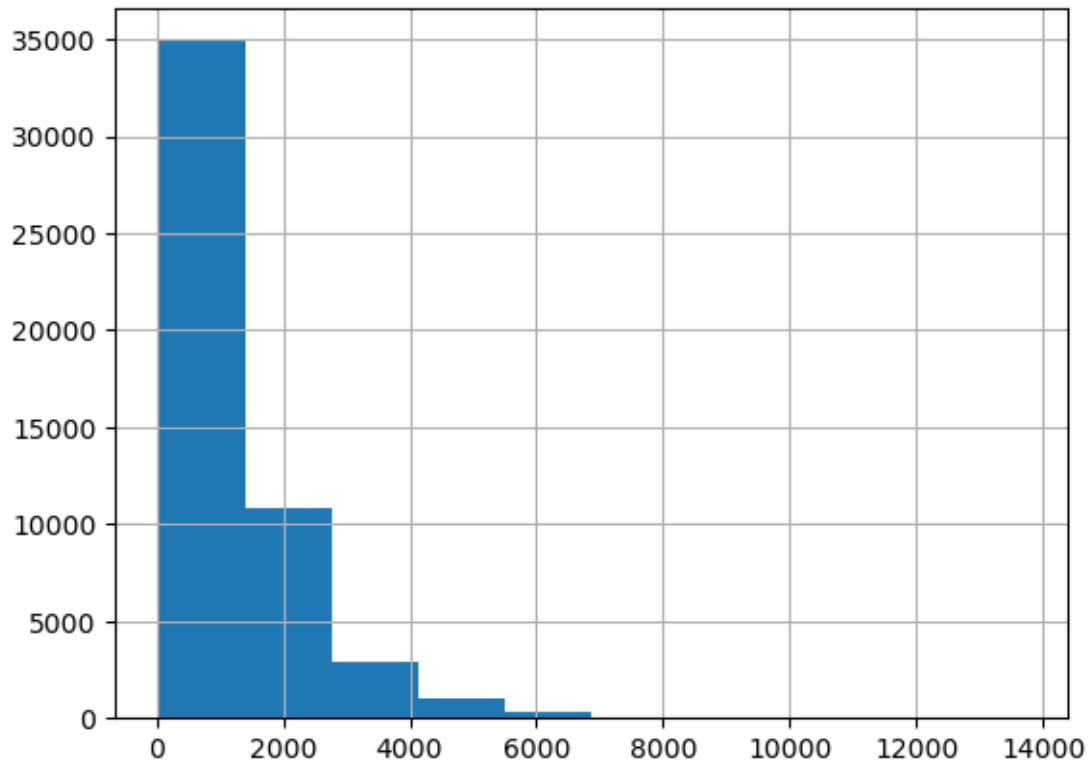
I would say this movie is not worth a rental, and hardly worth watching, however for the True Fan who needs to see all the movies, renting this movie is about the only way you'll see it - even the cable channels avoid this movie.

1
 Name: review, Length: 49582, dtype: int64

```
[15]: data['sentiment'].value_counts()
```

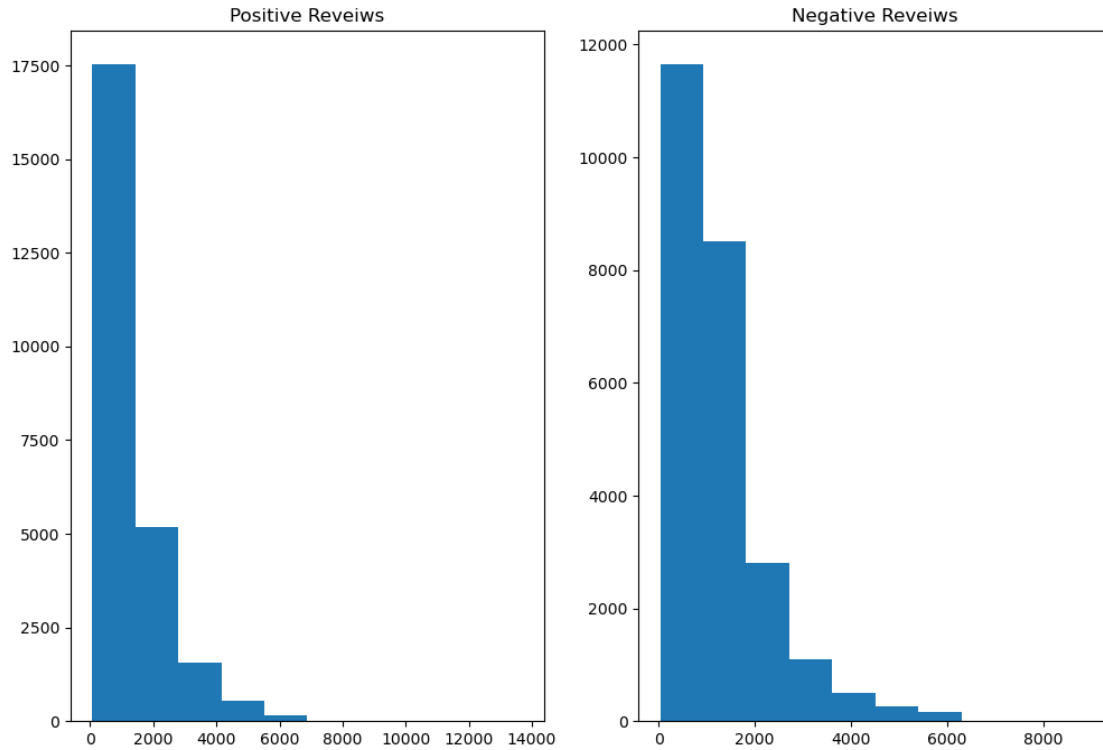
```
[15]: positive    25000
      negative    25000
      Name: sentiment, dtype: int64
```

```
[16]: data['review'].str.len().hist()
      plt.show()
```



```
[17]: # around 35,000 rows are having 0 to 1500 words and around 10,000+ rows are
      ↪having 1500 to 2500 words
```

```
[18]: fig,(ax1, ax2) = plt.subplots(1,2,figsize =[12,8])
      ax1.hist(data[data['sentiment'] == 'positive']['review'].str.len())
      ax1.set_title("Positive Reveiws")
      ax2.hist(data[data['sentiment'] == 'negative']['review'].str.len())
      ax2.set_title("Negative Reveiws")
      plt.show()
```



```
[19]: # word cloud
text = ' '.join(i for i in data [data['sentiment'] == 'positive']['review'])
wordcloud = WordCloud(background_color = 'white').generate(text)
plt.figure(figsize = (15,10))
plt.imshow(wordcloud, interpolation = 'bilinear')
plt.axis('off')
plt.title("WordCloud of Positive Reviews")
plt.show()
```

```
text = ' '.join(i for i in data[data['sentiment'] == 'negative']['review'])
wordcloud = WordCloud(background_color = 'white').generate(text)
plt.figure(figsize = (15,10))
plt.imshow(wordcloud, interpolation = 'bilinear')
plt.axis('off')
plt.title("WordCloud of Negative Reviews")
plt.show()
```

[illegible]


```
[21]: # let us clean the reviews column

data.rename(columns = {'review': 'text'}, inplace = True)
data.head()
```

```
[21]:                                     text sentiment
0  One of the other reviewers has mentioned that ... positive
1  A wonderful little production. <br /><br />The... positive
2  I thought this was a wonderful way to spend ti... positive
3  Basically there's a family where a little boy ... negative
4  Petter Mattei's "Love in the Time of Money" is... positive
```

```
[22]: def cleaning(text):
    # converting to lowercase, removing URL links, special characters,
    ↪punctuations...
    text = text.lower() # converting to lowercase
    text = re.sub('https?://\S+|www\.\S+', '', text) # removing URL links
    text = re.sub(r"\b\d+\b", "", text) # removing number
    text = re.sub('<.*?>+', '', text) # removing special characters,
    text = re.sub('%s' % re.escape(string.punctuation), '', text) #
    ↪punctuations
    text = re.sub('\n', '', text)
    text = re.sub('["'"]', '', text)

    #removing emoji:
    emoji_pattern = re.compile("[
        u"\U0001F600-\U0001F64F" # emoticons
        u"\U0001F300-\U0001F5FF" # symbols & pictographs
        u"\U0001F680-\U0001F6FF" # transport & map symbols
        u"\U0001F1E0-\U0001F1FF" # flags (iOS)
        u"\U00002702-\U000027B0" #Miscellaneous Symbols
        u"\U000024C2-\U0001F251" #Enclosed Alphanumeric
    ↪Supplement
        "]+", flags=re.UNICODE)
    text = emoji_pattern.sub(r'', text)

    # removing short form:

    text=re.sub("isn't",'is not',text)
    text=re.sub("he's",'he is',text)
    text=re.sub("wasn't",'was not',text)
    text=re.sub("there's",'there is',text)
    text=re.sub("couldn't",'could not',text)
    text=re.sub("won't",'will not',text)
    text=re.sub("they're",'they are',text)
    text=re.sub("she's",'she is',text)
    text=re.sub("There's",'there is',text)
```

```

text=re.sub("wouldn't",'would not',text)
text=re.sub("haven't",'have not',text)
text=re.sub("That's",'That is',text)
text=re.sub("you've",'you have',text)
text=re.sub("He's",'He is',text)
text=re.sub("what's",'what is',text)
text=re.sub("weren't",'were not',text)
text=re.sub("we're",'we are',text)
text=re.sub("hasn't",'has not',text)
text=re.sub("you'd",'you would',text)
text=re.sub("shouldn't",'should not',text)
text=re.sub("let's",'let us',text)
text=re.sub("they've",'they have',text)
text=re.sub("You'll",'You will',text)
text=re.sub("i'm",'i am',text)
text=re.sub("we've",'we have',text)
text=re.sub("it's",'it is',text)
text=re.sub("don't",'do not',text)
text=re.sub("that's",'that is',text)
text=re.sub("I'm",'I am',text)
text=re.sub("it's",'it is',text)
text=re.sub("she's",'she is',text)
text=re.sub("he's",'he is',text)
text=re.sub("I'm",'I am',text)
text=re.sub("I'd",'I did',text)
text=re.sub("he's",'he is',text)
text=re.sub("there's",'there is',text)
return text

```

```

[23]: data1 = data['text'].apply(cleaning)
      data1.head()

```

```

[23]: 0    one of the other reviewers has mentioned that ...
      1    a wonderful little production the filming tech...
      2    i thought this was a wonderful way to spend ti...
      3    basically theres a family where a little boy j...
      4    petter matteis love in the time of money is a ...
      Name: text, dtype: object

```

```

[24]: data1 = pd.DataFrame(data1)
      data1['sentiment'] = data['sentiment']
      data1.head()

```

```

[24]:                                     text sentiment
0  one of the other reviewers has mentioned that ...  positive
1  a wonderful little production the filming tech...  positive
2  i thought this was a wonderful way to spend ti...  positive

```

```

3 basically theres a family where a little boy j... negative
4 petter matteis love in the time of money is a ... positive

```

```

[25]: sw = stopwords.words('english')
      print(sw)

```

```

['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're",
"you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he',
'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's",
'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what',
'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am', 'is',
'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having',
'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or',
'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about',
'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above',
'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under',
'again', 'further', 'then', 'once', 'here', 'there', 'when', 'where', 'why',
'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some',
'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very',
's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now',
'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn',
"couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn',
"hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't",
'mustn', "mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn',
"shouldn't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn',
"wouldn't"]

```

```

[32]: data1['no_sw'] = data1['text'].apply(lambda x: " ".join([word for word in x.
↪split() if word not in (sw)]))
      data1.head()

```

```

[32]:                                     text sentiment \
0  one of the other reviewers has mentioned that ... positive
1  a wonderful little production the filming tech... positive
2  i thought this was a wonderful way to spend ti... positive
3  basically theres a family where a little boy j... negative
4  petter matteis love in the time of money is a ... positive

                                     no_sw
0  one reviewers mentioned watching oz episode yo...
1  wonderful little production filming technique ...
2  thought wonderful way spend time hot summer we...
3  basically theres family little boy jake thinks...
4  petter matteis love time money visually stunni...

```

```

[37]: from collections import Counter
      cnt = Counter()

```

```

for text in data1['no_sw'].values:
    for word in text.split():
        cnt[word] += 1
cnt.most_common(10)
temp = pd.DataFrame(cnt.most_common(10))
temp.columns=['word', 'count']
temp

```

```

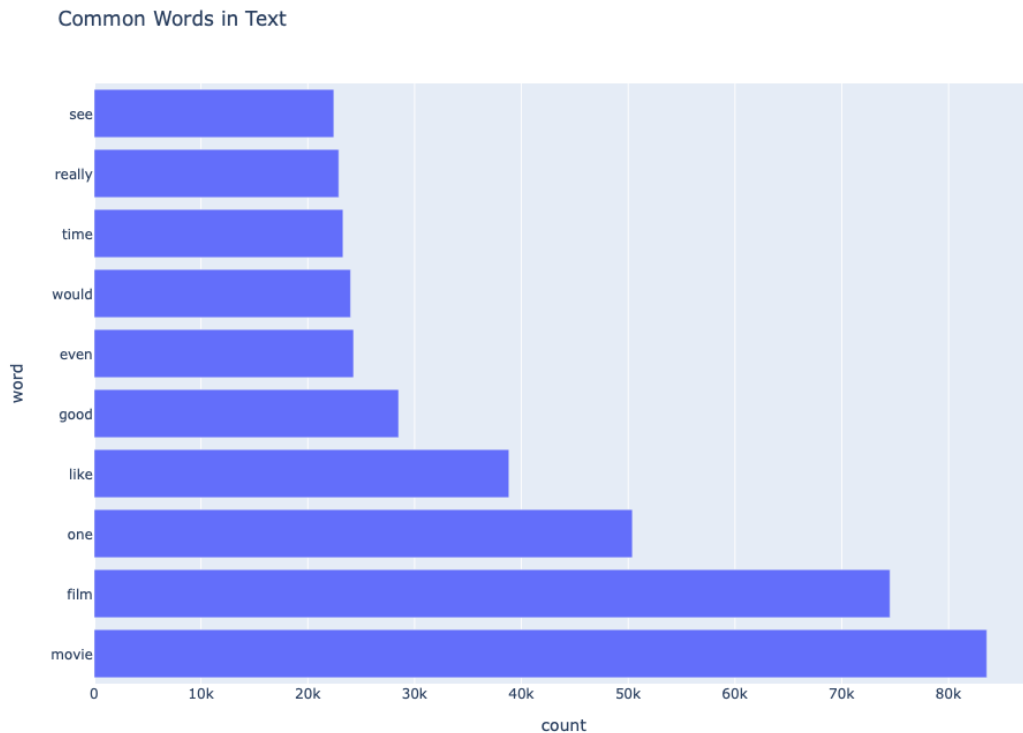
[37]:      word  count
0  movie  83576
1   film  74511
2    one  50392
3   like  38833
4   good  28502
5   even  24283
6  would  24001
7   time  23298
8 really  22900
9    see  22437

```

```

[39]: px.bar(temp, x = 'count', y = 'word', orientation = 'h', width = 700, height = 700, title = "Common Words in Text")

```



```
[43]: freq_words = set([w for (w,wc) in cnt.most_common(10)])
def remove_freq_words(text):
    return " ".join([word for word in str(text).split() if word not in
    ↪(freq_words)])
```

```
[45]: data1['wo_freq_words'] = data1['no_sw'].apply (lambda text :
    ↪remove_freq_words(text))
data1.head()
```

```
[45]:
```

	text	sentiment	\
0	one of the other reviewers has mentioned that ...	positive	
1	a wonderful little production the filming tech...	positive	
2	i thought this was a wonderful way to spend ti...	positive	
3	basically theres a family where a little boy j...	negative	
4	petter matteis love in the time of money is a ...	positive	

	no_sw	\
0	one reviewers mentioned watching oz episode yo...	
1	wonderful little production filming technique ...	
2	thought wonderful way spend time hot summer we...	
3	basically theres family little boy jake thinks...	
4	petter matteis love time money visually stunni...	

	wo_freq_words
0	reviewers mentioned watching oz episode youll ...
1	wonderful little production filming technique ...
2	thought wonderful way spend hot summer weekend...
3	basically theres family little boy jake thinks...
4	petter matteis love money visually stunning wa...

```
[70]: text = ' '.join(i for i in data1[data1['sentiment'] ==
    ↪'positive']['wo_freq_words'])
wordcloud = WordCloud(background_color = 'white').generate(text)
plt.figure(figsize = (15,10))
plt.imshow(wordcloud, interpolation = 'bilinear')
plt.axis('off')
plt.title("WordCloud of Positive Reviews")
plt.show()
```

```
text = ' '.join(i for i in data1 [data1['sentiment'] == 'negative']['wo_freq_words'])
wordcloud = WordCloud(background_color = 'white').generate(text)
plt.figure(figsize = (15,10))
plt.imshow(wordcloud, interpolation = 'bilinear')
plt.axis('off')
plt.title("WordCloud of Negative Reviews")
plt.show()
```

```
[47]: nltk.download('wordnet')
```

```
[nltk_data] Downloading package wordnet to  
[nltk_data]   /home/eec795a1-aada-4506-a501-  
[nltk_data]   7eee41ea48ce/nltk_data...  
[nltk_data]   Package wordnet is already up-to-date!
```

```
[47]: True
```

```
[53]: wordnet_lm = WordNetLemmatizer()  
data1['wordnet_lm'] = data1['wo_freq_words'].apply(wordnet_lm.lemmatize)  
data1.head()
```

```
[53]:
```

	text	sentiment	\
0	one of the other reviewers has mentioned that ...	positive	
1	a wonderful little production the filming tech...	positive	
2	i thought this was a wonderful way to spend ti...	positive	
3	basically theres a family where a little boy j...	negative	
4	petter matteis love in the time of money is a ...	positive	

	no_sw	\
0	one reviewers mentioned watching oz episode yo...	
1	wonderful little production filming technique ...	
2	thought wonderful way spend time hot summer we...	
3	basically theres family little boy jake thinks...	
4	petter matteis love time money visually stunni...	

	wo_freq_words	\
0	reviewers mentioned watching oz episode youll ...	
1	wonderful little production filming technique ...	
2	thought wonderful way spend hot summer weekend...	
3	basically theres family little boy jake thinks...	
4	petter matteis love money visually stunning wa...	

	wordnet_lm
0	reviewers mentioned watching oz episode youll ...
1	wonderful little production filming technique ...
2	thought wonderful way spend hot summer weekend...
3	basically theres family little boy jake thinks...
4	petter matteis love money visually stunning wa...

```
[65]: nb = data1.drop(columns = ['text', 'no_sw', 'wo_freq_words'])  
nb.columns = ['sentiment', 'wordnet_lm']  
nb.sentiment = [0 if each == 'negative' else 1 for each in nb.sentiment]  
nb.head()
```

```
[65]: sentiment                                wordnet_lm
0      1 reviewers mentioned watching oz episode youll ...
1      1 wonderful little production filming technique ...
2      1 thought wonderful way spend hot summer weekend...
3      0 basically theres family little boy jake thinks...
4      1 petter matteis love money visually stunning wa...
```

```
[67]: token_review = nb['wordnet_lm'].apply(lambda x: x.split())
token_review.head()
```

```
[67]: 0    [reviewers, mentioned, watching, oz, episode, ...
1    [wonderful, little, production, filming, techn...
2    [thought, wonderful, way, spend, hot, summer, ...
3    [basically, theres, family, little, boy, jake,...
4    [petter, matteis, love, money, visually, stunn...
Name: wordnet_lm, dtype: object
```

```
[ ]:
```