

Sentiment Analysis

Sentiment Analysis is a sub field of NLP that tries to identify the sentiment conveyed through a text. Sentiment analysis helps company in their decision making process.

Here we are performing sentiment analysis on movies reviews.

The 2,000 record IMDb movie review database is accessible through NLTK directly with

```
from nltk.corpus import movie_reviews
```

However, since we already have it in a tab-delimited file we'll use that instead.

In [1]:

```
#importing all the necessary libraries
import nltk
import pandas as pd
import string
from wordcloud import WordCloud, STOPWORDS
from nltk.corpus import stopwords
import matplotlib.pyplot as plt
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
```

In [2]:

```
#loading the dataset
df = pd.read_csv(r'TextFiles/moviereviews.tsv', sep = '\t')
df.head()
```

Out[2]:

	label	review
0	neg	how do films like mouse hunt get into theatres...
1	neg	some talented actresses are blessed with a dem...
2	pos	this has been an extraordinary year for austra...
3	pos	according to hollywood movies made in last few...
4	neg	my first press screening of 1998 and already i...

In [3]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 2 columns):
label    2000 non-null object
review   1965 non-null object
dtypes: object(2)
memory usage: 31.3+ KB
```

In [4]:

```
#checking for null values  
df.isnull().sum()
```

Out[4]:

```
label    0  
review   35  
dtype: int64
```

In [5]:

```
#dropping the NA values  
df.dropna(inplace=True)
```

In [6]:

```
df.isnull().sum()
```

Out[6]:

```
label    0  
review    0  
dtype: int64
```

In [7]:

```
#checking reviews having blank spaces(empty string) and removing them as well  
  
#starting with an empty list  
blanks= []  
  
for i, lb, rev in df.itertuples():#iterate over the dataframe  
    if type(rev) == str: #checking for string datatype  
        if rev.isspace(): #testing 'review' for whitespace  
            blanks.append(i) #adding the matching index no. to the list  
  
#dropping all the rows that has been appended into 'blanks'  
df.drop(blanks, inplace=True)
```

In [8]:

```
len(df)
```

Out[8]:

```
1938
```

Now we are left with 1938 rows in the dataset.

In [9]:

```
df['label'].value_counts()
```

Out[9]:

```
neg    969
pos    969
Name: label, dtype: int64
```

In [11]:

```
df.groupby('label').describe()
```

Out[11]:

	review			
	count	unique	top	freq
label				
neg	969	969	you know the plot : a dimwit with a shady past...	1
pos	969	969	on april 12th , 1912 , the most astonishing sh...	1

In [12]:

```
df['length'] = df['review'].apply(len)
```

In [13]:

```
df.length.describe()
```

Out[13]:

```
count    1938.000000
mean      3889.938080
std       1695.386131
min         92.000000
25%       2754.250000
50%       3631.000000
75%       4700.000000
max       15016.000000
Name: length, dtype: float64
```

In [14]:

```
df[df['length'] == 15016]['review'].iloc[0]
```

Out[14]:

'note : some may consider portions of the following text to be spoilers . \r\nbe forewarned . \r\nli
 ke its own opening shot , from out of the darkness boogie nights triumphantly explodes upon th
 e film scene with stunning authority and clarity , persuasively heralding its helmer paul thomas
 anderson as a bonafide major talent . \r\na sprawling epic set in the heart of the disco era and f
 ocused upon the ascension and decline of a troupe in the adult film industry , boogie nights is a
 n astonishingly ambitious feature film which is completely effective both as an utterly entertainin
 g joyride and as a strikingly impressive piece of filmmaking . \r\nthe film follows a soft-spoken y
 ounge man , eddie adams (mark wahlberg) from torrance , through his induction into the erotic
 picture business to stardom , and later , to a harrowingly precipitous drop . \r\nas boogie nights
 opens , he's toiling as a dishwasher in a trendy san fernando valley nightclub , and living at ho
 me under the thumb of a tyrannical mother who denigrates him at every opportunity . \r\nrecruit
 ed by fatherly jack horner (a rejuvenated burt reynolds) , a successful porn filmmaker , eddie a
 abandons his old life in favour of a new one as dirk diggler , pornstar . \r\nwholeheartedly embra
 ced by jack's troupe , including the maternal amber waves (julianne moore) , reed rothchild (j
 ohn c . reilly) , buck swope (don cheadle) , rollergirl (heather graham) , dirk and his " one sp
 ecial thing " are quickly propelled to the top of his field , fulfilling his mantra of being a " big brig
 ht shining star " . \r\nhis lack of discipline , however , incites a chain of circumstances which see
 s dirk and his friends unravel as a new decade is ushered in . \r\nthe element of boogie nights
 which literally commands attention is the bravura direction by mr . anderson , whose fascination
 and prowess with the camera is obvious . \r\nfrom the terrific tracking shot throughout a nightclu
 b which opens the film and introduces the players through to the showy sequence in the camer
 a follows a female partygoer diving into a pool , mr . anderson's accomplished command of the
 camera is without question . \r\nwhile novice directors often demonstrate an unfortunate lack of
 restraint in terms of flamboyant camerawork which ultimately proves to be aggravating , imperti
 nent and self-congratulatory (case in point : marc rocco's unforgettable perpetual-motion helmi
 ng of murder in the first) , in boogie nights every zoom or pan by mr . anderson , be it slow and
 deliberate or energetically sharp , is purposeful and potent . \r\nit's a rarity and a delight to disc
 over such a confidently assured work so early in a filmmaker's career -- boogie nights is only hi
 s second feature film , following up his debut feature from earlier this year , hard eight -- and alt
 hough it's somewhat derivative (his storytelling style is reminiscent of robert altman , while in t
 erms of visuals he's clearly influenced by martin scorsese) , at least he's borrowing from som
 e of the best , and it's entirely effective within the context of this film . \r\nwhile boogie nights is
 astonishingly polished from a technical perspective , in terms of sheer entertainment value it's
 also rollicking good fun , with a wickedly deadpan sense of humour . \r\nthe film perfectly captur
 es the essence of the disco period in terms of music , clothes , hairstyles , dance , idioms , and
 culture , aided immeasurably by impeccable production design by bob ziembicki and costume d
 esign by mark bridges , and boogie nights convincingly catapults the audience back to this rece
 nt , pre-aids-scare period in american history where uninhibited , free-wheeling lifestyles aboun
 ded . \r\nwhile much of the film's humour pertains to the characters' obliviousness of now-outd
 ated aspects of their lives then prevalent in society -- a prime example is part-time salesman bu
 ck swope's demo of a stereo system with an eight-track tape , and inviting a prospective custo
 mer to get 'freaky deaky' with the music -- it's elicited with an undercurrent not of ridicule , but
 of wistful reminiscence , and some the throwback elements in the film are strangely wondrous .
 \r\na giant ensemble disco dance production number in a nightclub is funny , to be sure , but od
 dly magical , and the sequence which lovingly explores eddie/dirk's bedroom , completely ador
 ned with posters and elements of iconic 1970s figures (farrah fawcett , cheryl tiegs , bruce lee)
 is breathtaking . \r\nably supported by a well-selected collection of period music (all of which p
 ropriately came straight from mr . anderson's personal collection) , there's a clear affection fo
 r the era in boogie nights . \r\nit's the period aspect and the playfully comic nostalgism of boogi
 e nights which has been played up in its marketing campaign in deference of the pornography e
 lement of the storyline ; the film's promotional trailer markedly avoids any explicit reference to t
 he adult film industry which unites its characters . \r\nafter the debacle with 1996's milos forma
 n film , the people vs . \r\nlarry flynt , another finely-crafted work which was knocked off the mar
 ket prematurely due to controversy regarding its pornography-related storyline , the new line m
 arketing team faced a unique challenge with boogie nights in terms of attempting to deemphasi

ze the relatively risqué erotic picture element of the story in order to increase salability to the general public . \r\nhowever , while boogie nights delves into the world of the adult film industry of two decades ago , it's hardly a lurid or seamy film -- given the subject matter , there's very little nudity , and virtually none of it can be construed as gratuitous -- and boogie nights smartly depicts erotic filmmaking not as a hotbed of carnality , but simply as a business . \r\nthe glimpses on the sets of jack horner's shoots reveal that while the films he makes are intended to titillate , there's a candid matter-of-factness in terms of sex ; his cast and crew are professionals who are merely doing a job , and it's refreshing that the obvious clichés of depravity and sinful lechery often linked to dismissive portraits of adult filmmaking , such as coercion or debauchery , are avoided . \r\nif anything , boogie nights is more interested in the process of adult filmmaking than the adult films themselves . \r\njack horner's dream is an honourable one -- he genuinely aspires to make a great film -- and to some degree boogie nights is reminiscent of ed wood (albeit lacking a certain wistful innocence) . \r\nin both films , the world of quickie low-budget filmmaking is explored , and both jack horner and the version of ed wood by tim burton , scott alexander , and larry karaszewski share similar mindsets -- jack's retort of " there are shadows in light , baby " to his cinematographer's complaint about poor set lighting echoes back to ed wood's " it's not about the little details , it's about the big picture ! " \r\nrant when it is pointed out that his graveyard set for plan 9 from outer space appears patently false . \r\nin boogie nights and ed wood , the aspirations of the respective filmmakers' prove to be bemusing , not in their intentions -- the real's nothing inherently silly in the goal of making an 'artistic' erotic picture -- but in the degree that they miss the mark . \r\nmuch like the humour lacing ed wood , the laughs involving jack horner are double-edged ; after viewing a hilarious excerpt from one of his latest films which hybridizes the james bond and porn genres , when jack breathes " this is the best work i've ever done " , it's genuinely funny , but given his utter sincerity , also possesses a tinge of sadness . \r\nmr . anderson is at this point clearly a greater talent as a director than a screenwriter . \r\nwhile boogie nights shows great inventiveness in staging , there really isn't any clever dialogue in the film , though a valid point certainly can be made that none of the film's characters particularly lend themselves to thoughtful verbal exchanges ; when one considers that the film's * protagonist * literally doesn't say a single insightful thing throughout the entire 152 minute running time , it's no wonder why the playfully inane " how much can you bench ? " \r\nbanter between dirk and reed constitutes one of the film's finest exchanges . \r\nand while mr . anderson creates a rich collection of fascinating characters , it's problematic that some of the threads given to the film's characters are simply stale material -- julianne moore's child custody battle is fairly clichéd stuff , better suited for a less risqué and uninspired tv-movie version of her character's story , and far less interesting than the thread dealing with her woefully hackneyed filmmaking ambitions . \r\nstill , even if he doesn't always know what to do with his characters , mr . anderson most definitely captures their essences , and his resolutely nonjudgemental affection for them is infectious -- misguided , even aimless , hopelessly naive , self-destructive , insufferable : i loved them just the same . \r\nin any case , it's hard to pick nits with a screenplay which contains one of the most audacious and inspired (and ultimately best-executed) scenes of the year : a scene where dirk diggler and reed rothchild are dragged along by their loose-cannon compadre to the home of drug baron rahad jackson (a mesmerizing alfred molina in an unforgettable performance , even if it's a small one) in a suicidal drug deal scam , only to find that the millionaire is a crazed eccentric (the part where he stops the deal to listen to the crescendo of night ranger's " sister christian " on his stereo system is sheer brilliance) flanked by an armed bodyguard . \r\nif this scene already wasn't tense enough with these elements , mr . anderson introduces a chinese boy that wanders through the scene , who for no discernable reason is tossing firecrackers around the room , pushing the scene to a new , surreal level of delightful giddiness . \r\nthere are some other interesting touches in boogie nights . \r\nmr . anderson recycles the use of long metaphorical chimes on the soundtrack from hard eight ; here , it accompanies the violent intercut sequences which depict the prostration of various characters to frightening effect . \r\nsimilarly , a painful scene where crew member scotty j (philip seymour hoffman) breaks down in tears in the driver's seat , sobbing " i'm so stupid " after being rebuffed by dirk after an awkward but heartfelt advance is strikingly played out exactly as a later scene with dirk in the driver's seat after a particularly tumultuous and traumatic evening . \r\nboogie nights is exceedingly well cast , and each of the actors rise to the occasion ; the ensemble cast is uniformly good . \r\nwhile most of the acting accolades will no doubt go to mr . wahlberg (fulfilling upon the sparks of promise he demonstrated in some of his earlier films) and mr . reynolds (terrific as the father figure and stable anchor of the film , ringing true in every respect) , a handful of other performances which might otherwise be overlooked deem mention . \r\nmr . reilly demonstrates here , as he did for brief glimpses in ulu grosbard's georgia , a genuine aptitude for comedy -- many of his scenes in

boogie nights are among the very funniest in the film . \r\ni'm not sure how much of his performance as reed was scripted and how much was done on the spot , but he plays the quintessential second banana character to a tee . \r\nmr . reilly is the best part of the big disco dance production number , and it's wickedly funny to watch his reed join in with dirk after the fact in threatening the new stud on the block during a macho altercation . \r\nworth the price of admission alone is the glimpse of him getting down in the recording studio while the dirk character is laying down a numbingly bad (and dead-on period ; it sure sounds like bad early 1980s to me) demo track " you got the touch " in a vain attempt at a musical career . \r\nmr . reilly plays his vacant character with utter conviction which is convincing and effective . \r\nms . graham , with her long dirty blonde hair , innocent demeanour , and huge hazel eyes , looks like a disney animated heroine come to life -- or in this case , a disney animated heroine who came to life , made a wrong turn , and ended up on the new line lot in a film focused on the late 70s / early 80s adult film industry . \r\nconsequently , it's a bit incongruous to see this cheery sweet-faced starlet whizzing about the set on her roller skates , but in a scene where she's lounging in a limousine prowling the streets , garish red lipstick messily smeared on , giving the camera a hilariously pouty come-hither look , all one can do is gulp . \r\nher role in the film sadly needs more fleshing out (err -- in a figurative sense) -- she's essentially a glorified extra -- but she does score impressively in her key scene which is unnerving in its sheer primal ferocity . \r\n(if only mr . anderson had filmed the scene so that we could actually see the expression of rage and anguish on her face , though !) \r\nand a cautionary tale for the strict and prudish : as a teen , ms . graham's parents forbade her to take a part in the twisted cult classic heathers ; since then , her most notable roles have been as a junkie (gus van sant's drugstore cowboy) and now a porn starlet . \r\n draw your own conclusions . \r\nperhaps the most quietly touching performance in boogie nights is that of mr . cheadle , who lends dignity to his role as a character in search of his identity , and whose quest is comically manifested by a series of changes in his sense of fashion . \r\nhis relationship with his cheerfully supportive new wife jessie st . vincent (an interesting melora walters) is wholly compelling has genuine sweetness , and although his general longing for acceptance from others is perceptible , it's no more so than in the bank scene where his loan application to open a stereo store is rejected due to his background in the porn industry . \r\nwhile the scenario is familiar , the wounded look in cheadle's eyes makes the scene unforgettable . \r\nluis guzman's most famous work is possibly in brian de palma's carlito's way , and here he plays a variation on the same character : maurice t . rodriguez is a hapless but good-natured nightclub owner who's a hanger-on to jack's group , ingratiating himself into the clique and constantly hounding jack , with little disguise made of his desperation , for a part in one of his erotic pictures . \r\njack's eventual acquiescence to maurice's request leads to one of boogie nights' biggest comic payoffs . \r\nboogie nights is a stunning achievement due to the virtuoso direction by mr . anderson , who immediately asserts his presence in the film industry as a presence of note with this work . \r\nstriking a fine balance between humour and drama , and exceptionally well-crafted , there's no doubt that boogie nights is among the very best films of the year . \r\n'

From the above output we can see that there are equal numbers of negative and positive reviews.

We will now clean the data by:

- Removing punctuations
- Removing all stopwords
- and Return a corpus of cleaned text

In [15]:

```
def text_clean(rev):

    # Check characters to see if they are in punctuation
    nopunc = [char for char in rev if char not in string.punctuation]

    # Join the characters again to form the string.
    nopunc = ''.join(nopunc)

    # Now just remove any stopwords
    nopunc = [w for w in nopunc.split() if w.lower() not in (stopwords.words('english'))]
    nopunc = ' '.join(nopunc)
    return nopunc
```

In [16]:

```
df["review"].head().apply(text_clean)
```

Out[16]:

```
0    films like mouse hunt get theatres isnt law so...
1    talented actresses blessed demonstrated wide a...
2    extraordinary year australian films shine scoo...
3    according hollywood movies made last decades l...
4    first press screening 1998 already ive gotten ...
Name: review, dtype: object
```

In [17]:

```
df['review'] = df['review'].apply(text_clean)
```

In [18]:

```
df.head()
```

Out[18]:

	label	review	length
0	neg	films like mouse hunt get theatres isnt law so...	2250
1	neg	talented actresses blessed demonstrated wide a...	3503
2	pos	extraordinary year australian films shine scoo...	3869
3	pos	according hollywood movies made last decades l...	5584
4	neg	first press screening 1998 already ive gotten ...	4327

In [19]:

```
df[df['length'] == 92]['review'].iloc[0]
```

Out[19]:

```
'film extraordinarily horrendous im going waste words'
```

In [21]:

```
df.describe()
```

Out[21]:

	length
count	1938.000000
mean	3889.938080
std	1695.386131
min	92.000000
25%	2754.250000
50%	3631.000000
75%	4700.000000
max	15016.000000

In [22]:

```
df['review'][0]
```

Out[22]:

'films like mouse hunt get theatres isnt law something diabolical load claptrap steven speilbergs dreamworks studio hollywood family fare deadly worst mouse hunt takes bare threads plot tries prop overacting flatout stupid slapstick makes comedies like jingle way look decent comparison writer adam rifkin director gore verbinski names chiefly responsible swill plot worth concerns tw o brothers nathan lane appalling lee evens inherit poorly run string factory seemingly worthless house eccentric father deciding check longabandoned house soon learn worth fortune set sellin g auction highest bidder battling every turn smart mouse happy rundown little abode wanting st ay way story alternates unfunny scenes brothers bickering inheritance endless action sequence s two take increasingly determined furry foe whatever promise film starts soon deteriorates bori ng dialogue terrible overacting increasingly uninspired slapstick becomes sound fury signifying nothing script becomes unspeakably bad best line poor lee evens utter another run rodent hate mouse oh cringe home alone ten times worse one touching scene early worth mentioning follo w mouse maze walls pipes arrives makeshift abode somewhere wall jumps tiny bed pulls make shift sheet snuggles sleep seemingly happy wanting left alone magical little moment otherwise soulless film message speilberg want dreamworks associated kind artistic credibility either give concerned mouse hunt swift kick arse hire decent writers directors kind rubbish'

NLTK's VADER Module

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER is an NLTK module that provides sentiment scores based on words used ("completely" boosts a score, while "slightly" reduces it), on capitalization & punctuation ("GREAT!!!!" is stronger than "great."), and negations (words like "isn't" and "doesn't" affect the outcome).

VADER's `SentimentIntensityAnalyzer()` takes in a string and returns a dictionary of scores in each of four categories:

- negative
- neutral
- positive
- compound

where the values for each falls between 0 and 1, and compound is the normalized score of the other three mentioned scores. Therefore a compound score of 0 would be completely neutral, a compound score above 0 indicates some sort of positive score and a compound score below 0 indicates negative score.

It doesn't require any training data but is constructed from a generalizable, valence-based, human-curated gold standard sentiment lexicon

In [23]:

```
from nltk.sentiment.vader import SentimentIntensityAnalyzer
```

In [24]:

```
sia = SentimentIntensityAnalyzer()
```

In [25]:

```
#creating three different columns depicting the VADER score, the compound score and the compound label(i.e.,
df['scores'] = df['review'].apply(lambda review: sia.polarity_scores(review))

df['compound'] = df['scores'].apply(lambda score: score['compound'])

df['comp_score'] = df['compound'].apply(lambda x: 'pos' if x>=0 else 'neg')
```

In [26]:

```
df.head()
```

Out[26]:

	label	review	length	scores	compound	comp_score
0	neg	films like mouse hunt get theatres isnt law so...	2250	{'neg': 0.167, 'neu': 0.665, 'pos': 0.168, 'co...	-0.4215	neg
1	neg	talented actresses blessed demonstrated wide a...	3503	{'neg': 0.178, 'neu': 0.668, 'pos': 0.154, 'co...	-0.8996	neg
2	pos	extraordinary year australian films shine scoo...	3869	{'neg': 0.116, 'neu': 0.642, 'pos': 0.241, 'co...	0.9952	pos
3	pos	according hollywood movies made last decades l...	5584	{'neg': 0.115, 'neu': 0.683, 'pos': 0.202, 'co...	0.9944	pos
4	neg	first press screening 1998 already ive gotten ...	4327	{'neg': 0.162, 'neu': 0.696, 'pos': 0.142, 'co...	-0.8947	neg

We'll now use scikit learn to determine how close VADER predicts the score as compared to our original data.

In [27]:

```
print(accuracy_score(df['label'], df['comp_score']))
```

0.6171310629514963

It looks like VADER couldn't analyse the movie reviews very accurately. This demonstrates one of the biggest challenges in sentiment analysis - understanding human semantics.

Splitting the data into training and test data

In [28]:

```
from sklearn.model_selection import train_test_split
```

In [29]:

```
X = df['review']  
y = df['label']
```

In [30]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 100)
```

Building pipeline to vectorize the data, then train and fit the model

In [31]:

```
from sklearn.pipeline import Pipeline  
from sklearn.feature_extraction.text import TfidfVectorizer  
from sklearn.naive_bayes import MultinomialNB  
from sklearn.svm import LinearSVC
```

Creating Pipeline with Naive Bayes classifier

In [32]:

```
text_clf_nb = Pipeline([('tfidf', TfidfVectorizer()),  
                        ('clf', MultinomialNB())])
```

In [33]:

```
text_clf_nb.fit(X_train, y_train)
```

Out[33]:

```
Pipeline(memory=None,
 steps=[('tfidf', TfidfVectorizer(analyzer='word', binary=False, decode_error='strict',
 dtype=<class 'numpy.float64'>, encoding='utf-8', input='content',
 lowercase=True, max_df=1.0, max_features=None, min_df=1,
 ngram_range=(1, 1), norm='l2', preprocessor=None, smooth_idf=True, ...rue,
 vocabulary=None)), ('clf', MultinomialNB(alpha=1.0, class_prior=None, fit_prior=True))])
```

In [34]:

```
predict1 = text_clf_nb.predict(X_test)
```

Creating pipeline with SVM classifier

In [35]:

```
text_clf_svm = Pipeline([('tfidf', TfidfVectorizer()),
                          ('clf', LinearSVC())])
```

In [36]:

```
text_clf_svm.fit(X_train, y_train)
```

Out[36]:

```
Pipeline(memory=None,
 steps=[('tfidf', TfidfVectorizer(analyzer='word', binary=False, decode_error='strict',
 dtype=<class 'numpy.float64'>, encoding='utf-8', input='content',
 lowercase=True, max_df=1.0, max_features=None, min_df=1,
 ngram_range=(1, 1), norm='l2', preprocessor=None, smooth_idf=True, ...ax_iter=1000,
 multi_class='ovr', penalty='l2', random_state=None, tol=0.0001,
 verbose=0))])
```

In [37]:

```
predict2 = text_clf_svm.predict(X_test)
```

Evaluating the Model

In [38]:

```
print("~MultinomialNB~\n")
print(confusion_matrix(y_test, predict1))
print(classification_report(y_test, predict1))
print('Accuracy Score:', accuracy_score(y_test, predict1))
```

~MultinomialNB~

```
[[163 31]
 [ 48 146]]
      precision    recall  f1-score   support

 neg     0.77     0.84     0.80     194
 pos     0.82     0.75     0.79     194

 micro avg     0.80     0.80     0.80     388
 macro avg     0.80     0.80     0.80     388
 weighted avg     0.80     0.80     0.80     388
```

Accuracy Score: 0.7963917525773195

In [39]:

```
print("~SVM~\n")
print(confusion_matrix(y_test, predict2))
print(classification_report(y_test, predict2))
print('Accuracy Score:', accuracy_score(y_test, predict2))
```

~SVM~

```
[[163 31]
 [ 31 163]]
      precision    recall  f1-score   support

 neg     0.84     0.84     0.84     194
 pos     0.84     0.84     0.84     194

 micro avg     0.84     0.84     0.84     388
 macro avg     0.84     0.84     0.84     388
 weighted avg     0.84     0.84     0.84     388
```

Accuracy Score: 0.8402061855670103

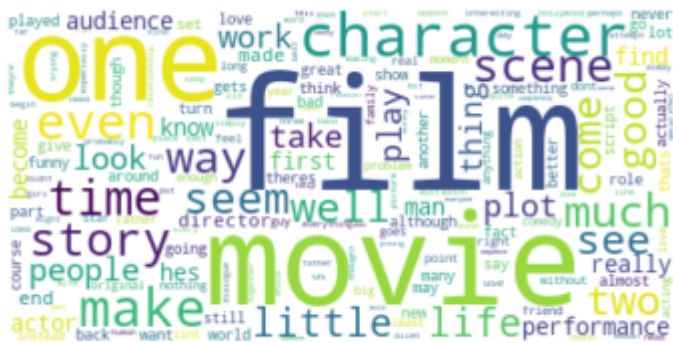
It seems that the SVM classifier has given better accuracy of 84% as compared to Naive Bayes with accuracy of 79%

Generating WordCloud to visualize the frequency of words that has been

In [40]:

```
text = " ".join(rev for rev in df.review)
```

```
stopwords = set(STOPWORDS)
# Generate a word cloud image
wordcloud = WordCloud(stopwords=stopwords, background_color="white").generate(text)
# Display the generated image:
# the matplotlib way:
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```



```
review = "This movie was pathetic"
print(text_clf.nb.predict([review]))
```

['neg']

In []: