

Sentiments Analysis using Naive Bayes

May 15, 2020

Data Analysis and Visualization

```
[1]: # importing require libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline

import seaborn as sns
```

```
[2]: # load dataset
df = pd.read_csv('sentiment.csv')
df.head()
```

```
[2]:
```

	Phrase	label
0	Spider-man is better than any summer blockbust...	Positive
1	of good sense	Positive
2	, then knock yourself out and enjoy the big sc...	Negative
3	So could young romantics out on a date .	Positive
4	humour	Positive

```
[3]: # shape of the dataset
df.shape
```

```
[3]: (2800, 2)
```

```
[4]: # checking how many positive or negative values are there
df.label.value_counts()
```

```
[4]: Positive    1800
Negative    1000
Name: label, dtype: int64
```

```
[5]: # checking the null values
df.isnull().sum()
```

```
[5]: Phrase    0
label        0
```

dtype: int64

```
[6]: # finding the unique values from the phrase
unique_words = list(df['Phrase'].str.split(' ', expand=True).stack().unique())

print(unique_words)
```

```
['Spider-man', 'is', 'better', 'than', 'any', 'summer', 'blockbuster', 'we',
'had', 'to', 'endure', 'last', ',', 'and', 'hopefully', 'sets', 'the', 'tone',
'for', 'a', 'of', 'good', 'stuff', '.', 'sense', 'then', 'knock', 'yourself',
'out', 'enjoy', 'big', 'screen', 'postcard', 'that', 'self-glorified', 'Martin',
'Lawrence', 'lovefest', 'So', 'could', 'young', 'romantics', 'on', 'date',
'humour', 'You', 'bet', 'there', 'it', "'s", 'what', 'makes', 'this', 'rather',
'convoluted', 'journey', 'worth', 'taking', 'tries', 'more', 'Screenwriter',
'Chris', 'ver', 'Weil', 'directing', 'debut', 'good-natured', 'never', 'dull',
'but', 'mood', 'in', 'which', 'an', 'ominous', 'pervasive', 'unknown', 'threat',
'lurks', 'just', 'below', 'proceedings', 'adds', 'almost', 'constant',
'mindset', 'suspense', 'centers', 'midlife', 'crisis', 'hate', 'putting',
'together', 'movies', 'particular', 'value', 'or', 'merit', 'sensual', 'funny',
'Some', 'computer', 'animation', 'handsome', 'right', 'questions', 'solid',
'--', 'as', 'if', 'director', 'trying', 'dupe', 'viewer', 'into', 'all', 'Very',
'Important', 'simply', 'because', 'movie', 'ugly', 'look', 'at', 'not',
'Hollywood', 'product', 'thriller', 'once', 'too', 'often', 'war', 'One',
'year', 'most', 'weirdly', 'engaging', 'unpredictable', 'character', 'pieces',
'brilliant', 'they', 'do', "n't", 'fit', 'well', 'keep', 'interested', 'proves',
'sometimes', 'dreams', 'youth', 'should', 'remain', 'hack', 'Fire', 'nostalgic',
'twisty', 'yarn', 'will', 'them', 'guessing', 'brings', 'proper', 'conviction',
'his', 'role', '-LRB-', 'Jason', 'Bourne', '-RRB-', 'delicious', 'pulpiness',
'its', 'lurid', 'fiction', 'stylized', 'with', 'touch', 'John', 'Woo', 'bullet',
'ballet', 'you', "'re", 'entirely', 'unprepared', 'clever', 'credits', 'roll',
'ploughing', 'same', 'furrow', 'gently', 'swaying', 'back', 'forth', 'cradles',
'characters', 'veiling', 'tension', 'beneath', 'otherwise', 'tender',
'movements', "'", 'laugh', 'therapy', '"', 'whose', 'view', 'America',
'history', 'awkwardness', 'human', 'life', 'generous', 'deep', 'like', 'work',
'artist', 'who', 'tired', 'fighting', 'fights', 'weight', 'world', 'shoulders',
'playing', 'narrative', 'form', 'pair', 'fascinating', 'performances',
'Phillip', 'Noyce', 'actors', 'cinematographer', 'Christopher', 'Doyle',
'understand', 'delicate', 'forcefulness', 'Greene', 'prose', 'impossible',
'care', 'about', 'very', 'impression', 'evoke', 'memories', 'emotions', 'are',
'anything', 'drag', 'reek', 'script', 'A', 'well-made', 'humane', 'children',
'film', 'truest', 'plays', 'dramatic', 'Invigorating', 'surreal', 'resonant',
'rainbow', 'emotion', 'might', 'Vulgar', 'bilingual', 'charmer', 'can',
'honestly', 'describe', 'looking', 'sounding', 'feeling', 'no', 'other',
'recent', '```', 'Austin', 'Powers', 'Goldmember', "''", 'has', 'silly',
'entertainment', 'enough', 'laughs', 'sustain', 'interest', 'end', 'aims', 'be',
'uplifting', 'moving', 'occasional', 'smiles', 'surprisingly', 'achievement',
'really', 'special', 'walk', 'woods', 'show', 'us', 'time', 'courage', 'ideas',
```

'technical', 'proficiency', 'evil', 'monstrous', 'lunatic', 'strong', 'themes',
 'familial', 'ties', 'well-intentioned', 'effort', 'still', 'burdened', 'by',
 'actor', 'sneaky', 'feel', 'several', 'old', 'unintentional', 'parallels',
 'inadvertently', 'humorous', 'so', 'intimate', 'psychologically', 'self-
 revealing', 'passes', 'sex', 'cheap', 'hysterics', 'This', 'some', 'honest',
 'insight', 'black-and-white', 'inspires', 'cliche', 'happy', 'ending',
 'Swedish', 'fillm', 'modern', 'city', 'first', 'computer-generated', 'feature',
 'cartoon', 'have', 'been', 'funnier', 'released', 'outtakes', 'theatrically',
 'used', 'bonus', 'DVD', 'visual', 'charm', 'would', 'hard', 'sitting',
 'through', 'one', 'spirals', 'downward', 'inoffensive', 'goose', 'fun', 'over-
 indulgent', 'tirade', 'love', 'story', 'sanguine', 'title', 'More',
 'vaudeville', 'well-constructed', 'Takes', 'clunky', 'TV-movie', 'approach',
 'detailing', 'chapter', 'celebrated', 'Irish', 'playwright', 'poet', 'drinker',
 'confident', 'filmmaking', 'much-needed', 'levity', 'disappointed', 'lets',
 'brush', 'up', 'against', 'humanity', 'psycho', 'without', 'making', 'him',
 'less', 'the-night', 'pat', 'storylines', 'precious', 'circumstances',
 'beautiful', 'stars', 'success', 'wonderful', 'Madonna', 'gives', 'her', 'best',
 'performance', 'since', 'Abel', 'Ferrara', 'beaten', 'pulp', 'Dangerous',
 'Game', 'Eight', 'Legged', 'Freaks', 'perfectly', 'entertaining', 'diversion',
 'contemplation', 'fused', 'eerie', 'atmosphere', 'deceit', 'insanely',
 'violent', 'graphic', 'Starts', 'mediocre', 'thuds', 'bottom', 'pool',
 'utterly', 'incompetent', 'conclusion', 'An', 'unsettling', 'memorable',
 'cinematic', 'experience', 'coming', 'down', 'off', 'Miramax', 'shelves',
 'after', 'couple', 'aborted', 'attempts', 'sweet', 'At', 'when',
 'commercialism', 'squeezed', 'whatever', 'idealism', 'American', 'moviemaking',
 'ever', 'Godfrey', 'Reggio', 'career', 'shines', 'lonely', 'beacon', 'tryingly',
 'Keenly', 'observed', 'refreshingly', 'natural', 'mainstream', 'audiences',
 'rarely', 'seen', 'indication', 'Got', 'depress', 'itself', 'Superbly',
 'Filmmaker', 'Tian', 'Zhuangzhuang', 'triumphantly', 'returns', 'visually',
 'masterful', 'quiet', 'power', 'captures', 'innocence', 'budding', 'demons',
 'within', 'wallflower', 'created', 'monster', 'did', 'know', 'how', 'handle',
 'If', 'horror', 'primary', 'goal', 'frighten', 'disturb', 'They', 'works',
 'spectacularly', '...', 'shiver-inducing', 'nerve-rattling', 'ride', 'give',
 'accurate', 'thought', 'right-thinking', 'films', 'revel', 'splendor', 'Feels',
 'Even', 'fans', 'Ismail', 'Merchant', 'I', 'suspect', 'indie', 'powerful',
 'people', 'fully', 'does', 'leave', 'much', 'ham-fisted', 'introspective',
 'independent', 'thoughtful', 'art', 'ethics', 'cost', 'moral', 'compromise',
 'B', 'Perhaps', 'call', 'refreshing', 'endless', 'assault', 'embarrassingly',
 'jokes', 'Reeks', 'rot', 'home', 'French', 'shocker', 'usual', 'bad', 'boy',
 'weirdo', 'team', 'offbeat', 'humor', 'amusing', 'warmest', 'relic', 'from',
 'bygone', 'era', 'convolutions', 'plausible', 'bizarre', 'Although', 'fairly',
 'involving', 'far', 'goes', 'having', 'fresh', 'say', 'growing', 'Catholic',
 'uncluttered', 'gem', 'leaps', 'over', 'national', 'boundaries', 'celebrates',
 'universal', 'nature', 'waxes', 'poetic', 'our', 'taste', 'amazing',
 'slapstick', 'instrument', 'different', 'own', 'yet', 'instantly',
 'recognizable', 'heroes', 'try', 'avoid', 'succeed', 'few', 'cool', 'seems',
 'aware', 'coolness', 'little', 'recommend', 'Snow', 'Dogs', 'Waking', 'Up',
 'Reno', 'case', 'letting', 'sleeping', 'dogs', 'lie', 'The', 'rehashes',

'capped', 'pointless', 'extremes', 'hellish', 'conditions', 'feels', 'pretty',
'execution', 'pomposity', 'solving', 'problem', 'distract', 'solution',
'another', 'poorly', 'delivered', 'smug', 'self-satisfaction', 'usually',
'associated', 'private', 'schools', 'alone', 'force', 'scratch', 'hole', 'your',
'head', 'experimental', 'storytelling', 'losing', 'stooping', 'base',
'melodrama', 'decent', 'material', 'music', 'lot', 'richer', 'ones', 'action',
'screenwriters', 'come', 'their', 'loosely-connected', 'string', 'mythic',
'constructs', 'relays', 'points', 'lectures', 'confrontations', 'beyond',
'comprehension', 'even', 'were', 'Moonlight', 'Mile', 'judgment', 'damned',
'Hartley', 'emphasizes', 'well-wrought', 'omits', 'needless', 'chase', 'scenes',
'swordfights', 'revenge', 'unfolds', 'all-enveloping', 'Provides', 'kind',
'need', 'comedies', 'spiritual', 'riveting', 'set', 'comedy', 'equivalent',
'song', 'finally', 'found', 'vent', 'Importance', 'Being', 'Earnest', 'thick',
'wit', 'reading', 'Bartlett', 'Familiar', 'Quotations', 'gritty', 'major',
'weird', 'Masterpiece', 'Theater', 'sketch', 'neither', 'concentrating',
'elements', 'revealing', 'alienation', 'among', 'culture', 'sadly', 'hostile',
'odds', 'recklessness', 'retaliation', 'When', 'rich', 'luscious', 'nuances',
'empty', 'intrigue', 'betrayal', 'murder', 'light', 'Errol', 'Morris',
'focusing', 'eccentricity', 'failing', 'ultimately', 'make', 'something',
'bigger', 'scrapbook', 'oddballs', 'attract', 'upscale', 'hungry', 'quality',
'corny', 'examination', 'affecting', 'portrait', 'screwed-up', 'man', 'dared',
'mess', 'Written', 'flatly', 'Mr.', 'Polanski', 'element', 'here', 'zingers',
'barely', 'defensible', 'sexual', 'violence', 'Inventive', 'intoxikatingly',
'sexy', 'self-indulgent', 'maddening', 'example', 'lush', 'predictable',
'reassuring', 'manner', 'beautifully', 'sung', 'holiday', 'carol', 'vapid',
'exercise', 'Best', 'delivers', 'perfect', 'get', 'breezy', 'blend', 'banality',
'big-screen', 'remakes', 'Avengers', 'Wild', 'West', 'crafted', 'wisecracking',
'Mystery', 'Science', '3000', 'guys', 'keeps', 'IMAX', 'short', 'X', 'blab',
'what-if', 'photographed', 'staged', 'Mendes', 'series', 'vulgar',
'forgettably', 'dark', 'real', 'reason', 'made', 'Mamet', 'House', 'Games',
'fall', 'Heist', 'prevention', 'place', 'blame', 'baseball', 'macabre', 'where',
'religious', 'civic', 'virtues', 'hold', 'society', 'tatters', 'Both', 'lead',
'Oscar-size', 'actually', 'filmmakers', 'want', 'nothing', 'else', 'way',
'Whenever', 'threatens', 'bogged', 'earnest', 'dramaturgy', 'fairy', 'tale',
'comes', 'renowned', 'Indian', 'allows', 'Americans', 'preposterous', 'moments',
'seeking', 'Like', 'sizzle', 'news', 'exotic', 'creatures', 'overboard',
'loony', 'melodramatic', 'denouement', 'high', 'school', 'swimming',
'substitutes', 'bathtub', 'evanescent', 'seamless', 'sumptuous', 'stream',
'consciousness', 'only', 'quest', 'president', 'single-handedly', 'turned',
'plane', 'full', 'hard-bitten', 'cynical', 'journalists', 'was', 'essentially',
'campaign', 'extended', 'publicity', 'department', 'graced', 'Poor', 'editing',
'bluescreen', 'ultra-cheesy', 'dialogue', 'complex', 'those', 'crass',
'contrived', 'sequels', 'fails', 'second-guess', 'affection', 'original',
'position', 'Thanks', 'largely', 'emotionally', 'least', 'Wiseman', 'fatal',
'error', 'flash', 'rock', 'videos', 'rip-roaring', 're-creations', 'famous',
'gratuitous', 'distractions', 'chuckles', 'gag', 'positively', 'thrilling',
'combination', 'ethnography', 'good-bad', 'Eerily', 'moment', 'high-concept',
'candy-coat', 'spirituality', 'definite', 'room', 'improvement', 'Robert',

'Harmon', 'less-is-more', 'bump-in', '-', 'chills', 'greatest', 'triumph',
 'keeping', 'creepy', 'crawlies', 'hidden', 'shadows', 'hews', 'carries',
 'effortlessly', 'darkness', 'wondering', 'lives', 'offers', 'opportunities',
 'Cinema', 'Paradiso', 'find', 'new', 'interesting', '"d"', 'probably', 'turn',
 'convinced', 'already', 'employ', 'kids', 'owe', 'favors', 'parents',
 'companion', 'Its', 'generic', 'villains', 'lack', 'accents', 'cinema',
 'wondrously', 'gifted', 'artists', 'Arthur', 'Schnitzler', 'Reigen', 'Sy',
 'open-faced', 'smiling', 'madmen', 'difficult', 'worthy', 'airless', 'shell',
 'games', 'latest', 'eccentric', 'super-wealthy', 'megalomaniac', 'bent',
 'domination', 'destruction', 'romantic', 'Proves', 'formulas', 'messaging',
 'serves', 'workable', 'primer', 'region', 'Manages', 'please', 'intended',
 'audience', 'placing', 'coma-like', 'state', 'occurs', '7', 'times', 'during',
 'Windtalkers', 'serious-minded', 'modest', 'winning', 'misses', 'prisoner',
 'Smoke', 'Signals', 'also', 'imbued', 'paralyzed', 'stoner', 'midnight',
 'flick', 'sci-fi', 'deconstruction', 'cliches', 'considerable', 'dash',
 'running', 'around', 'screaming', 'death', 'style', 'sincere', 'veers', 'comic',
 'course', 'It', 'gets', 'doubt', 'asks', 'country', 'serious', 'ways',
 'extravagant', 'chance', 'distort', 'perspective', 'throw', 'path',
 'incoherence', 'despite', 'impressed', 'upon', 'injects', 'freshness', 'acting-
 workshop', 'exercises', 'Saddam', 'Hussein', 'In', 'cut', 'rhapsodizes', 'add',
 'bleak', 'Trouble', '25', 'minutes', 'gratitude', 'basic', 'comedic', 'likes',
 'franchise', 'years', 'small', 'easily', 'overshadowed', 'predictability',
 'esoteric', 'musings', 'philosophy', 'emotional', 'connections', 'absorbing',
 'Scarcely', 'expense', 'paid', 'pay', 'see', 'shows', 'slice', 'production',
 'values', 'Malcolm', 'D.', 'Lee', 'writer', 'Ridley', 'Downright',
 'transparent', 'rewrite', 'designed', 'garner', 'cooler', 'PG-13', 'rating',
 'welcome', 'relief', 'stop', 'mom', 'amiable', 'degree', 'randomness',
 'achieved', 'lottery', 'drawing', 'plotting', 'details', 'promenade', 'clad',
 'bodies', 'Myrtle', 'Beach', 'S.C.', 'adrenaline', 'jolt', 'sudden', 'lunch',
 'rush', 'diner', 'unfocused', 'hero', 'open', 'devolves', 'ends', 'satire',
 'mind', 'adolescent', 'demanding', 'regular', 'shocks', 'bouts', 'Mostly',
 'Martha', 'mostly', 'unsurprising', 'And', 'appreciate', 'one-sided', 'theme',
 'predecessors', 'sensational', 'true-crime', 'theatres', 'many', 'things',
 'delectable', 'everything', 'expect', 'number', 'tumbleweeds', 'blowing',
 'terms', 'bites', 'chew', 'linking', 'massacre', 'Armenians', '1915',
 'relationships', 'present', 'rut', 'life-changing', 'encounters', 'bagatelle',
 'nearly', 'epic', 'proportions', 'Violent', 'superior', 'stale', 'uninspired',
 'toxic', 'intoxicating', 'may', 'believe', 'puts', 'battle', 'wills',
 'revelatory', 'narcissistic', 'morality', 'While', 'About', 'Boy', 'provide',
 'enjoyable', '100', 'theater', 'Swimfan', 'Fatal', 'Attraction', 'brilliantly',
 'played', 'deeply', 'highest', 'wears', 'Mark', 'Pellington', 'pop', 'kooky',
 'overeager', 'spooky', 'subtly', 'myth', 'Mediocre', '"m"', 'ready', 'go',
 'U.N.', 'ask', 'permission', 'preemptive', 'strike', 'occasionally',
 'horrifying', 'inspiring', 'expected', 'grace', 'Miyazaki', 'storytellers',
 'achieving', 'cliched', 'perverse', 'escapism', 'Charming', 'apart',
 'reporting', 'flickering', 'reminders', 'bind', 'Mood', 'Beautifully', 'miss',
 'joyous', 'unless', 'considers', 'source', 'hilarity', 'sure', 'filmmaker',
 'disagree', 'point', 'chronicle', 'smart', 'provocative', 'drama', ':',

'mention', 'company', 'overexposed', 'waste', 'gorgeously', 'atmospheric',
 'meditation', 'Beating', 'sad', 'professional', 'conventional', 'comedy-drama',
 'post-camp', 'old-time', 'Run', 'Do', 'judge', 'soon', 'Bloody', 'grieving',
 'process', 'transcends', 'boy-meets-girl', 'posturing', 'typical', 'stories',
 'outnumber', 'hits', 'three-to-one', 'appropriate', 'compelling',
 'increasingly', 'diverse', 'seeing', 'great', 'deal', 'though', 'Sillier',
 'cuter', 'shorter', 'remember', 'bit', 'heavy', 'handed', 'message',
 'recommended', 'viewing', 'acting', 'Spielberg', 'realization', 'near-future',
 'Horrendously', 'amateurish', 'plainly', 'incomprehensible', 'recognized',
 'mildly', 'Changing', 'Lanes', 'mystical', 'Robinson', 'web', 'matches', 'page-
 turning', 'frenzy', 'Clancy', 'creates', 'heavy-handed', 'screenplay', 'Too',
 'clumsy', 'key', 'splash', 'absolute', 'joy', 'poetry', 'common', 'through-
 line', 'pinnacle', 'going', 'house', 'party', 'watching', 'host', 'defend',
 'himself', 'frothing', 'ex-girlfriend', 'sophomoric', 'romp', 'happen', 'Has',
 'rare', 'capability', 'soothe', 'break', 'heart', 'single', 'stroke', 'tragedy',
 'record', 'tenacious', 'fighter', 'victim', 'Well-done', 'supernatural',
 'catch', 'lacks', 'quick', 'Steven', 'Schindler', 'List', 'Williams',
 'developments', 'processed', '60', 'intellectual', 'lector', 'auteur',
 'injuries', 'Parents', 'beware', ';', 'hopeful', 'maintaining', 'appearance',
 'clinical', 'objectivity', 'start', 'speeds', 'Burr', 'Steers', 'knows',
 'territory', 'particularly', 'Witty', 'Those', 'saved', 'aid', 'teeth', 'hurt',
 'mawkish', 'self-parody', 'hardly', 'exist', 'responsible', 'generates', 'plot',
 'What', 'begins', 'evolves', 'Imax', 'chimps', 'lots', 'blown', 'size',
 'latently', 'gay', 'scattered', 'bogus', 'There', 'grenade', 'rooted',
 'undergoing', 'cares', '?', 'Judith', 'Zaza', 'bedroom', 'sequence', 'bulk',
 'wrong', 'heartfelt', 'middle', 'dedication', 'Aggressive', 'self-
 glorification', 'manipulative', 'whitewash', 'consistently', 'unimaginative',
 'sophisticated', 'literate', 'such', 'pictures', 'standard', 'procedure',
 'prince', 'fellow', 'survival', 'awkwardly', 'magic', 'realism', 'meaty',
 'face', 'thing', 'tackled', 'subject', 'drawn', 'while', 'peppering', 'pages',
 'Analyze', 'That', 'popcorn', 'Every', 'Day', 'plodding', 'Vincent', 'Gallo',
 'finish', 'meaning', 'phrase', 'rambling', 'living', 'mug', 'shots',
 'impeccable', 'pedigree', 'mongrel', 'pep', 'Uneven', 'nice', 'prove', 'potent',
 'hell-jaunt', 'purists', 'Tolstoy', 'groupies', 'tricky', 'satisfying', 'sharp',
 'homage', 'Them', 'Tarantula', 'low', '10', 'compressed', 'Tells', 'terrific',
 '10th-grade', 'learning', 'tool', 'smartly', 'fact', 'Disney', 'scrape', 'me',
 'successful', 'dragons', 'Reign', 'Shainberg', 'adaptation', 'gander', 'amuses',
 'none', 'amounts', 'Sleek', 'arty', 'daring', 'verve', 'outpaces',
 'contemporaries', 'weak', 'wounds', 'inspired', 'indecipherable',
 'complications', 'he', 'revealed', 'before', 'abandoned', 'consoled',
 'depiction', 'depression', 'non-techies', 'Though', 'rent', 'parable',
 'inexorably', 'rote', 'sentimentality', 'Hampered', '"ve", 'feature-length',
 'antics', 'pleasure', 'cracker', 'barrel', 'reaction', 'subjects', 'furiously',
 'Genuinely', 'unnerving', 'mesmerizing', '', 'disguised', 'images', 'attracts',
 'unnamed', 'substitutable', 'forces', 'serve', 'terror', 'll", 'admit', 'my',
 'complaint', 'Amid', 'shock', 'curiosity', 'factors', 'actress', 'idealistic',
 'kid', 'chooses', 'champion', 'cause', 'pretentiousness', 'Damon', 'Wilde',
 'surprising', 'lazy', 'budget', 'B-movie', 'thrillers', '1950s', '"60s',

'Campanella', 'directed', 'expedience', 'roiling', 'documentary', 'Sunday',
 'flair', 'slight', 'poem', 'eyes', 'put', 'hairs', 'chest', 'told', 'liking',
 'Ice', 'Age', 'given', 'day', 'comprehensible', 'Dummies', 'guide', 'Daughter',
 'frame', 'Intriguing', 'downright', 'completely', 'emphasis', 'James', 'Eric',
 'Horton', 'Peter', 'O'Fallon', 'stirring', 'dog', 'Winds', 'being', 'both',
 'surge', 'swirling', 'rapids', 'leap', 'rouses', 'mainland', 'setting',
 'Transforms', 'Shakespeare', 'deepest', 'tragedies', 'proud', 'Ki-Deok',
 'After', 'woman', 'run-of-the-mill', 'ruthless', 'social', 'order', 'governs',
 'college', 'cliques', 'flame-like', 'trembling', 'progression', 'two',
 'skittish', 'New', 'York', 'middle-agers', 'stumble', 'relationship',
 'struggle', 'fears', 'foibles', 'Mary', 'Gaitskill', 'harrowing', 'improved',
 'destination', 'mixed', 'results', 'escapades', 'demonstrating', 'adage', 'wo',
 'safe', 'Plays', 'vignettes', 'clips', 'Pow', 'Swimming', 'radical', 'Just',
 'everyone', 'involved', 'coasting', 'But', 'followed', 'runaway', 'Full',
 'Monty', 'inept', 'Jones', 'Evokes', 'David', 'raunch', 'favor', 'gags', 'rely',
 'strength', 'cleverness', 'Report', 'card', 'Does', 'live', 'exalted',
 'tagline', 'sidekicks', 'Devos', 'Oedekerker', 'childhood', 'dream', 'martial-
 arts', 'lose', 'above', 'Shakespearean', 'Less', 'efforts', 'toward', 'closure',
 'highlight', 'Spalding', 'Gray', 'teen', 'gross-out', 'Love', 'Hong', 'Kong',
 'From', 'Danang', 'reveals', 'scientific', 'law', 'discerned', 'producers',
 'heed', 'Hatfield', 'Hicks', 'oddest', 'couples', 'becomes', 'study', 'gambles',
 'publishing', 'offering', 'exists', 'political', 'ramifications', 'glacial',
 'pacing', 'fan', 'fantasia', 'Monte', 'Cristo', 'war-ravaged', 'land',
 'unlikely', 'Sarah', 'Harrison', 'keen', 'insights', 'parapsychological',
 'phenomena', 'soulful', 'Jolting', 'under', 'skin', 'freakish', 'seem', 'saw',
 'TV', '300', 'Russian', 'mighty', 'Ivan', 'swear', 'rest', 'game', 'hooked',
 'As', 'Storytelling', 'version', 'Quiet', 'creature-feature', 'Shyamalan',
 'boilerplate', 'punch', 'lines', 'unfortunately', 'double-cross', 'Swinging',
 'hobby', 'Nothing', 'various', 'eventually', 'All', 'arch', 'cold-blooded',
 'Claude', 'Chabrol', 'camera', 'Entertains', 'Kim', 'tenderness', 'takes',
 'generate', 'heat', 'juicy', 'soap', 'opera', 'Cho', 'creating', 'navigates',
 'fast', 'fade', '!', 'blaxploitation', 'spoof', 'downplays', 'opposed',
 'extent', 'outrageousness', 'fest', 'Using', 'stock', 'scruffy', 'slip',
 'convincing', 'dizzily', 'gorgeous', 'nor', 'Narratively', 'drive-by',
 'tentative', 'familiar', 'anti-feminist', 'equation', '=', 'misery', 'ambition',
 'whole', 'staring', 'Kendall', 'structure', 'Offers', 'Initial', 'strangeness',
 'Q', 'Quirky', 'Now', 'former', 'Gong', 'Show', 'addict', 'enthusiasm', 'Not',
 'scares', 'cross', 'chemicals', 'bunch', 'totally', 'unexpected', 'directions',
 'Green', 'arriving', 'core', 'Tartakovsky', 'powers', 'memory', 'monstrous',
 'penance', 'cowering', 'begging', 'feet', 'Giannini', 'Narc', 'strikes',
 'defiantly', 'retro', 'chord', 'willingness', 'experiment', 'explosions',
 'early', 'trusted', 'sort', 'toothless', 'book', 'Favor', 'document', 'Comes',
 'across', 'Highly', 'Contrived', 'depressing', 'ruthlessly', 'pained',
 'depraved', 'wound', 'Scherfig', 'five', 'writers', 'He', 'Fresnadillo',
 'Videodrome', '100-minute', 'Something', 'co-writer', 'Ed', 'Decter', 'senses',
 'mean-spirited', 'Before', 'broken', 'outweighs', 'loss', 'witness', 'visible',
 'His', 'preach', 'exclusively', 'converted', 'Writer\\director', 'Kung',
 'Seeing', 'momentum', 'remains', 'undeterminable', 'roughshod', 'sensibilities',

```
'makers']
```

```
[7]: # count the unique word from phrase
unique_word_count = df.Phrase.str.split(' ', expand=True).stack().value_counts()

print(unique_word_count)
```

```
the          995
,            817
a            740
and          702
of           701
...
His           1
Scherfig      1
spoof         1
bunch         1
safe          1
Length: 1979, dtype: int64
```

Here, the most frequent word is **the** and total number of unique words is 1979.

```
[8]: # counting the words in each row
df['WordCount'] = df['Phrase'].str.split().str.len()

df.head()
```

```
[8]:
```

	Phrase	label	WordCount
0	Spider-man is better than any summer blockbust...	Positive	27
1	of good sense	Positive	3
2	, then knock yourself out and enjoy the big sc...	Negative	19
3	So could young romantics out on a date .	Positive	9
4	humour	Positive	1

```
[9]: # checking if there is empty Phrase or not
df[df['WordCount'] == 0]
```

```
[9]:
```

	Phrase	label	WordCount
981		Negative	0

```
[10]: # dropping the index which has no phrase
df.drop(df[df['WordCount'] == 0].index, inplace=True)

df.shape
```

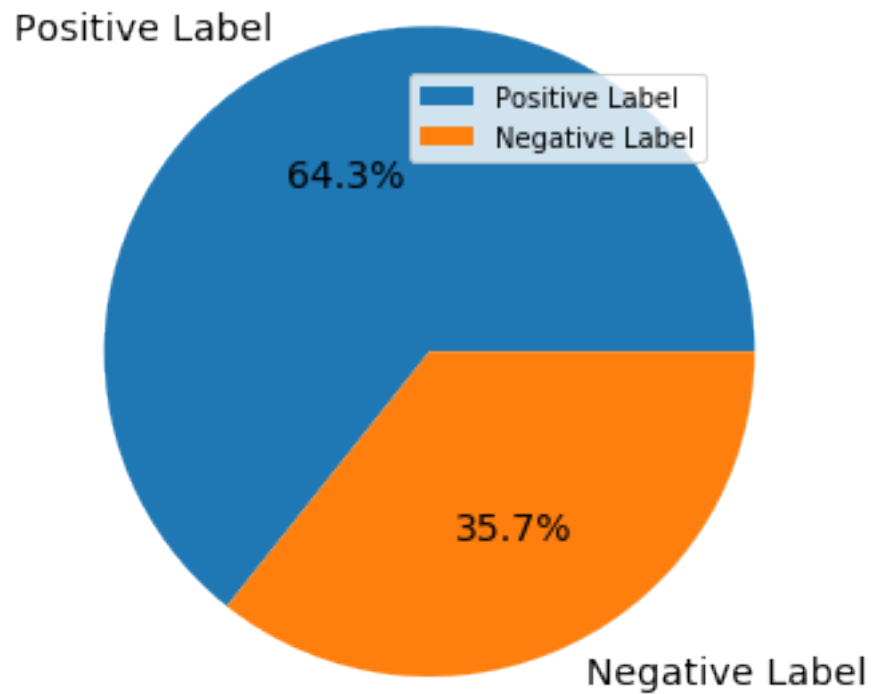
```
[10]: (2799, 3)
```



```
[11]: # plotting the pie chart
pos_total = len(df[df['label'] == 'Positive'])
neg_total = len(df[df['label'] == 'Negative'])

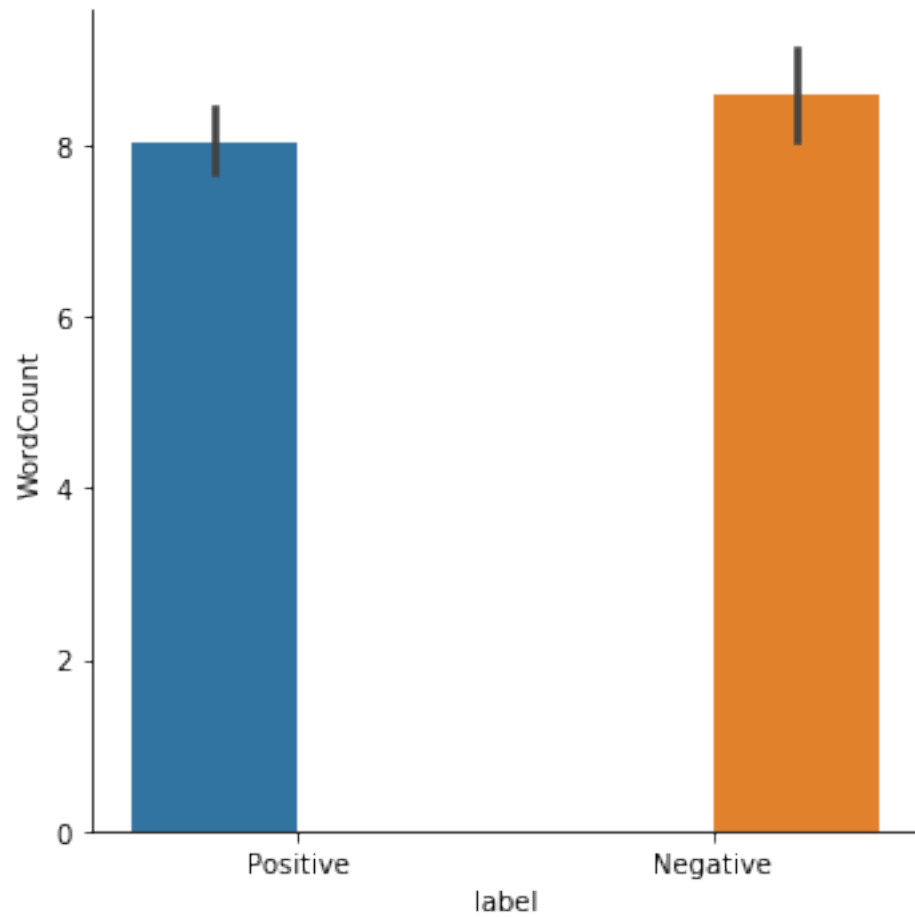
textprops = {'fontsize': 14}
plt.pie(x=[pos_total, neg_total], labels=['Positive Label', 'Negative Label'],
        textprops=textprops, autopct='%0.1f%%', radius=1.4)

plt.legend()
plt.show()
```

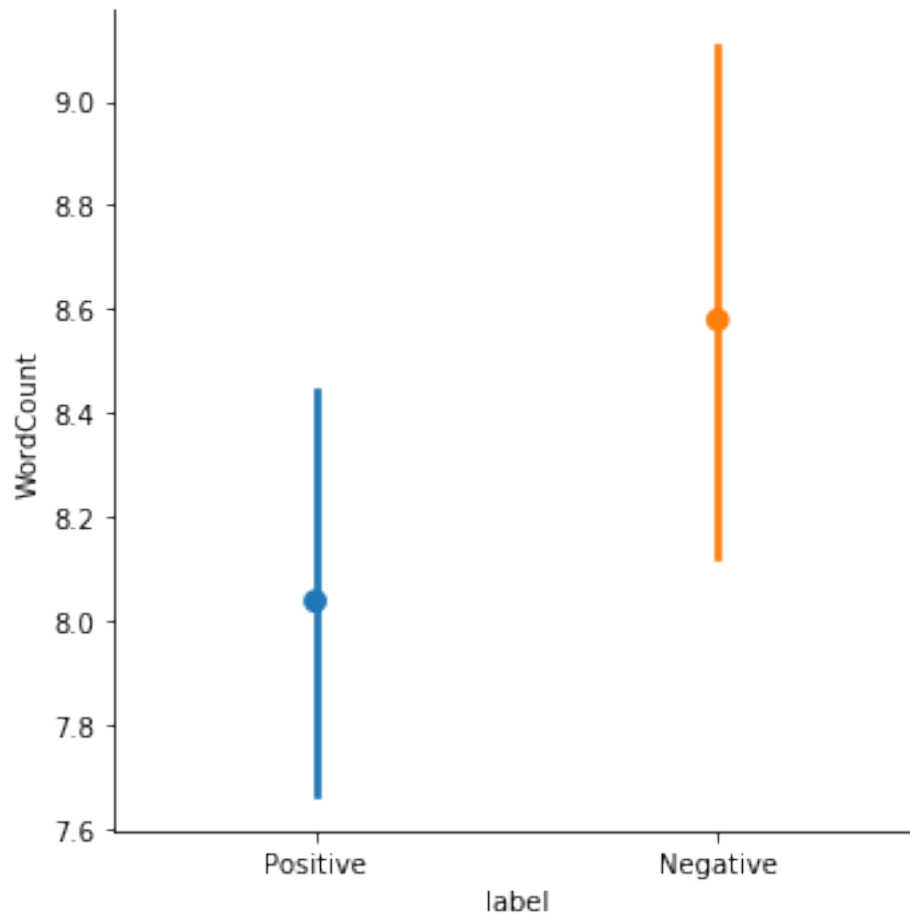


```
[12]: # average wordcount per plot

sns.catplot(x='label', y='WordCount', data=df, kind='bar', hue='label')
plt.show()
```



```
[13]: # plotting the categorical data  
  
sns.catplot(x='label', y='WordCount', data=df, kind='point', hue='label')  
plt.show()
```



The above plots shows that the **Negative** label has slightly more word count than the **Positive** word count in general.

```
[14]: # Data cleaning
new_df = df.copy()

[15]: # removing (, . !)symbol from the dataset using regular expression
import re

def clean_phrase(dataset):
    clean_data = re.sub("[^a-z\s']+", " ", dataset, flags = re.IGNORECASE)

    return clean_data

[16]: # clean data column of phrase
new_df['Phrase'] = new_df['Phrase'].apply(clean_phrase)
new_df.head()
```

```
[16]:
```

	Phrase	label	WordCount
0	Spider man is better than any summer blockbust...	Positive	27
1	of good sense	Positive	3
2	then knock yourself out and enjoy the big sc...	Negative	19
3	So could young romantics out on a date	Positive	9
4	humour	Positive	1

```
[17]: # new word_count after removing (, . !) from dataframe
new_df['word_count'] = new_df['Phrase'].str.split().str.len()
new_df.head()
```

```
[17]:
```

	Phrase	label	WordCount	\
0	Spider man is better than any summer blockbust...	Positive	27	
1	of good sense	Positive	3	
2	then knock yourself out and enjoy the big sc...	Negative	19	
3	So could young romantics out on a date	Positive	9	
4	humour	Positive	1	

	word_count
0	25
1	3
2	18
3	8
4	1

```
[18]: # changing label to 1 and 0
def encoding_label(dataset):
    if dataset == 'Positive':
        dataset = 1
    elif dataset == 'Negative':
        dataset = 0

    return dataset

new_df['label'] = new_df['label'].apply(encoding_label)
new_df.head()
```

```
[18]:
```

	Phrase	label	WordCount	\
0	Spider man is better than any summer blockbust...	1	27	
1	of good sense	1	3	
2	then knock yourself out and enjoy the big sc...	0	19	
3	So could young romantics out on a date	1	9	
4	humour	1	1	

	word_count
0	25
1	3

```
2      18
3      8
4      1
```

Calculating of prior, likelihood and posterior

```
[19]: # splitting number for train and test dataset
total_len = len(new_df)
train_len = int(0.6 * total_len)
test_len = int(0.4 * total_len)

print(total_len, train_len, test_len)
```

```
2799 1679 1119
```

```
[20]: # splitting dataframe
train_df = new_df[: train_len]
test_df = new_df[train_len: ]

print(len(train_df), len(test_df))
```

```
1679 1120
```

```
[21]: # creating unique vocab
all_vocab = list()

for i in train_df['Phrase'].tolist():
    all_vocab.extend(i.split(" "))

unique_vocab = list(set(all_vocab))

print('total_vocab : {}, unique_vocab : {}'.format(len(all_vocab),
    ↪len(unique_vocab)))
```

```
total_vocab : 15134, unique_vocab : 1842
```

```
[22]: # calculating the class priors
class_priors = train_df['label'].value_counts(normalize = True)
class_priors
```

```
[22]: 1    0.646814
0    0.353186
Name: label, dtype: float64
```

```
[23]: # getting class 0 (Negative) and 1 (Positive) words from phrase
class_0_words = []

for sentence in train_df[train_df['label'] == 0]['Phrase']:
```

```

word_list = sentence.lower().split(" ")

new_word_list = []
for w in word_list:
    if w != '':
        new_word_list.append(w)

class_0_words.extend(new_word_list)

```

```

[24]: class_1_words = []

for sentence in train_df[train_df['label'] == 1]['Phrase']:
    word_list = sentence.lower().split(" ")

    new_word_list = []
    for w in word_list:
        if w != '':
            new_word_list.append(w)

    class_1_words.extend(new_word_list)

```

```

[25]: # class word dictionary
class_words = {
    0: class_0_words,
    1: class_1_words
}

```

Smoothing

Counting occurrence of word and finding the word probabilities for respective classes and performing laplacian smoothing.

```

[26]: # function
def count_occur(word, word_list):
    count = 0

    for w in word_list:
        if word == w:
            count += 1

    return count

def word_probs_with_laplacian_smoothing(class_0_words, class_1_words, alpha, unique_vocab):
    word_probs = {
        0: {},
        1: {}
    }

```

```

    #Laplacian Smoothing
    for word in unique_vocab:
        word_probs[0][word] = ((count_occur(word, class_0_words) + alpha) /
↪(len(class_0_words) + alpha * len(unique_vocab)))
        word_probs[1][word] = ((count_occur(word, class_1_words) + alpha) /
↪(len(class_1_words) + alpha * len(unique_vocab)))

    return word_probs

```

Prediction

```

[27]: def prediction(alpha, dataset):
        word_probs = word_probs_with_laplacian_smoothing(class_0_words,
↪class_1_words, alpha, unique_vocab)
        predicted_class = []

        for phrase in dataset:
            phraseArr = phrase.split(" ")
            prob_0_given_string = 1
            prob_1_given_string = 1

            for w in phraseArr:
                try:
                    prob_0_given_string = prob_0_given_string*word_probs[0][w]
                    prob_1_given_string = prob_1_given_string*word_probs[1][w]
                except Exception:
                    w_prob_class_0 = ((alpha/
↪len(class_0_words)+alpha*len(unique_vocab)))
                    prob_0_given_string = prob_0_given_string*w_prob_class_0
                    w_prob_class_1 = ((alpha/
↪len(class_1_words)+alpha*len(unique_vocab)))
                    prob_1_given_string = prob_1_given_string*w_prob_class_1
                    prob_0_given_string *= class_priors[0]
                    prob_1_given_string *= class_priors[1]

            if prob_1_given_string > prob_0_given_string:
                predicted_class.append(1)
            else:
                predicted_class.append(0)

        return predicted_class

```

```

[28]: predictions = prediction(1, test_df["Phrase"])

```

```

[29]: # confusion matrix
def get_confusion_matrix(actual, predicted):

```

```

confusion_matrix = np.zeros((2,2))
for x, y in zip(actual, predicted):
    confusion_matrix[x, y] += 1
return confusion_matrix

```

```

[30]: confusion_matrix = get_confusion_matrix(test_df['label'], predictions)

print('Confusion Matrix:\n', confusion_matrix)

```

```

Confusion Matrix:
[[233. 173.]
 [ 29. 685.]]

```

```

[31]: # accuracy
accuracy=(confusion_matrix[0][0] + confusion_matrix[1][1]) / np.
    ↳sum(confusion_matrix)

print('Accuracy: ', accuracy)

```

```

Accuracy:  0.8196428571428571

```

```

[32]: # recall
recall=(confusion_matrix[0][0]) / (confusion_matrix[0][0] +↳
    ↳confusion_matrix[0][1])

print('Recall: ', recall)

```

```

Recall:  0.5738916256157636

```

```

[33]: # precision
precision=(confusion_matrix[0][0]) / (confusion_matrix[0][0] +↳
    ↳confusion_matrix[1][0])

print('Precision: ', precision)

```

```

Precision:  0.8893129770992366

```

```

[34]: # f1 score
f1_score=(2 * recall * precision) / (recall + precision)

print('F1 Score: ', f1_score)

```

```

F1 Score:  0.6976047904191617

```

ROC curve

```

[35]: tpr = []
      fpr = []

```



```

for alpha in np.arange(1,100,1):
    predictions = prediction(alpha,test_df["Phrase"])
    confusion_matrix = get_confusion_matrix(test_df["label"], predictions)

    true_pos_rate = (confusion_matrix[0][0]) /
    ↳(confusion_matrix[0][0]+confusion_matrix[0][1])
    false_pos_rate = (confusion_matrix[1][0]) /
    ↳(confusion_matrix[1][0]+confusion_matrix[1][1])

    tpr.append(true_pos_rate)
    fpr.append(false_pos_rate)

```

```

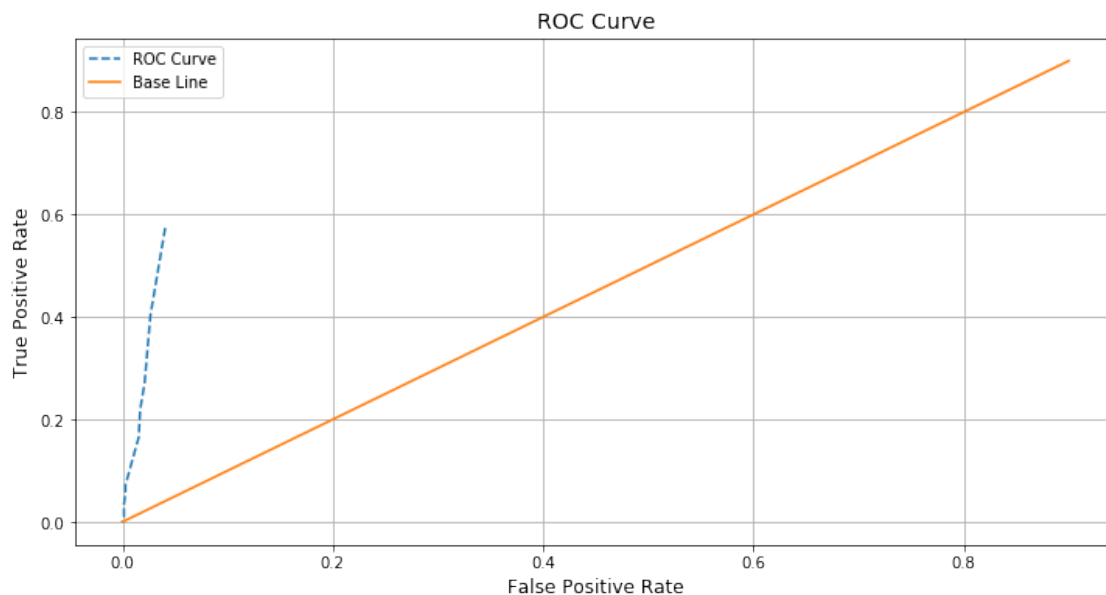
[36]: # plotting ROC curve
plt.figure(figsize = (12, 6))
base = np.arange(0,1,0.1)

plt.plot(fpr, tpr, linestyle='--', label='ROC Curve')
plt.plot(base, base, label='Base Line')

plt.title('ROC Curve', fontsize = 14)
plt.xlabel('False Positive Rate', fontsize = 12)
plt.ylabel('True Positive Rate', fontsize = 12)

plt.legend()
plt.grid()
plt.show()

```



[]: