NORTHWESTERN UNIVERSITY | SCHOOL OF PROFESSIONAL STUDIES

Aspect-based Sentiment Analysis

Yining Feng

March 16, 2021

Abstract

One key task of fine-grained sentiment analysis of movie reviews is to extract movie aspects or features that film critics have expressed opinions on. In general, opinion mining is quite context-sensitive, and, at a coarser granularity, quite domain dependent. This study introduces an aspect-based approach for opinion mining, which uses the ontology structure as an essential part of the feature extraction process, by taking account the relations between entities. In addition, this study investigates Support Vector Machines, Naïve Bayes, Logistic Regression, and Bi-LSTM classification models. The focus of the study is the comparison of these classifiers by evaluating their classification accuracy, based on the size of training data sets, and hyperparameter tuning. In this study, movie reviews retrieved from the **imdb.com** website in the early 2000s by Bo Pang and Lillian Lee as well as movie reviews collected from 12 popular and authoritative sources for film ratings and critical reviews are analyzed, respectively.

1. Introduction & Problem Statement

The main goal is to improve aspect-level opinion mining by employing ontology. The project will be broken down into five steps/stages: part-of-speech tagging, name entity recognition, ontology development, text classification, and sentiment scoring (positive, neutral, negative).

Named Entity Recognition and Classification (NERC) is a process of recognizing information units like names, including person, organization and location names, and numeric expressions including time and date from movie review documents. The goal is to develop practical and domain-independent techniques in order to detect named entities with high accuracy automatically. In this study, the `spaCy` Python library for Natural Language Processing will be exclusively deployed for the named entity recognition task. Subsequently, Support Vector Machines, Naïve Bayes, Logistic Regression, and Bi-LSTM classification models are used to solve text classification tasks. These machine learning methods can train themselves and define relationships between features of the objects. Thus, this study intends to employ these methods to get the best classification accuracy using movie-review data collected from various websites.

For sentiment scoring step, polarity measurement will be developed based on a lexicon of tagged positive and negative sentiment terms which are used to quantify positive/negative sentiment. In this part, a sentiment analysis framework called VADER sentiment will be used as it provides a readily interpretable positive and negative polarity value for a set of "affective" terms. This study will focus on two tasks for sentiment analysis:

(1) convert the polarity obtained from the polarity identification step to more precise one by analyzing contextual features, such as "negation" rules.

(2) with the help of ontology again, make it possible to compute the polarities of the nodes through the hierarchy relationship.

2. Research Design and Modeling Method

2.1 Data Description and Processing

The move review polarity dataset for training classification models is a collection of cleaned up movie reviews retrieved from the imdb.com website in the early 2000s by Bo Pang and Lillian Lee. The reviews were collected and made available as part of their research on natural language processing. The reviews were originally released in 2002, but an updated and cleaned up version was released in 2004, referred to as v2.0. The dataset is comprised of 1,000 positive and 1,000 negative movie reviews drawn from an archive of the rec.arts.movies.reviews newsgroup hosted at IMDB, with a cap of 20 reviews per author (312 authors total) per category (Pang & Lee, 2004).

The new movie review dataset for model evaluation is comprised of 42 movie review documents collected from New York Times, Washington Post, RogerEbert.com, The Seattle Times, Vanity Fair, IGN, Hollywood Reporter, Breathe Dream Go, The Guardian, Variety, NPR, and Cinemajam. Each movie review document is trimmed to 500 words and contains the movie title, plot summary, release date, actors /actresses, directors, movie review texts and so on. All 42 movie review texts would be cleaned and pre-processed by using special natural language processing features, such as: *segmentation* by separating each single word with punctuation or white space, removing all stop words, such as *a* and *the*, or by making all capital letters a lower case; *stemming* by reducing words to their base or root forms; *term frequency* by counting the frequency of words which helps identify how important a word is to a document in a corpus; *word embedding* is transformation of words to an array of numeric values of semantic or contextual information that computer can understand. The resulting dataset includes 42 unlabeled movies reviews trimmed from 21,896 terms to 9,835 terms.

2.2 Methodology

VADER Sentiment Scoring

Valence Aware Dictionary and Sentiment Reasoner (VADER) is a lexical sentiment classifier, and it is used to do initial sentiment labeling of each of the 42 new movie reviews. A sentiment lexicon is a lexicon where the words have been annotated with semantic scores, often between -1 and 1. VADER sentiment can also aggregate sentiment scores from individual words into sentence scores. The support for sentence sentiment also takes into account booster words (e.g. *very* in very happy) and negation words (e.g. *not* in not happy). For sentiment scoring, the `compound` score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). In this study, any movie review with the value of `compound` score ≥ 0.05 is classified as a positive review, whereas any movie review with the value of `compound` score ≤ - 0.05 is classified as a negative review. The rest of movie reviews each with the value of `compound` score between - 0.05 and 0.05 are classified as neutral reviews. This metric provides a normalized, weighted composite score, which can be used as a single unidimensional measure of sentiment for a given sentence.

Movie Review Ontology

Semantic modeling and Ontology are the most common techniques for inferring and modeling contextual information from users' data. Expressing context values using ontology is advantageous because ontology can reveal various characteristics and properties of the context. Ontologies represent the systematically classification of items. The concepts used in the ontology proposed in this study are categorized and related to movie, genre, themes, and geographic region. The nouns are considered as objects and the verbs as object properties. These classes, objects and object properties' information are used to build the ontology model. The Stanford's Protégé software is used to build the ontology model. Class, object and object property are identified as entity, individual, and object property in the ontology model, respectively. The relations between classes, objects and object properties were derived manually as per the human understanding of a sentence. The ontology would use Web Ontology Language (OWL) for structural specifications and 42 movie review texts as the data source.

In an ontology, concepts are classes and subclasses of a domain, object properties represent relationships among various objects, and data properties represent attributes of the objects. To build a movie ontology, first of all, all the entities included in movie reviews are identified (Figure 4), such as persons, places, and concepts, etc. The top class is “owl: Thing”. There are three subclasses: Concept, Movie Thing, and NewsStation. While instances of “Concept” are derived from tokens and phrases extraced from movie reviews, there is only one instance of “NewsStation” since all my movie reviews are downloaded from New York Times. Next, the subclass “Movie Thing” is broken down into five subclasses: Genre, Movie, Person, Place, and Review. All 42 movies fall under the subclass “Movie”. Each movie is assigned with 1 data property (“hasReleaseYear”) and 8 object properties: “hasReviewby”, “hasActor”, “happensinPlace”, “hasBiographicalReference”, “hasCharacter”, “hasConcept”, “hasDirector”, “hasGenre”. On the other hand, 6 categories are created for the “Person” subclass: Actor, Author, Character, Director, BiographicalReference, Role. Each instance of “Person” is assigned with one object property “hasRoleas”. In addition, each instance of “Person” (except those in the “Role” category representing occupations) is assigned with two data properties “hasName” and “hasGender”. Similarly, the “Place” subclass, which refers to the narrative place in each movie, is divided into “City” and “Country” categories. Meanwhile, 15 instances are created for the subclass “Genre”: Action, Adventure, Comedy, Crime, Documentary, Drama, Family, Fantasy, History, Political, Romance, Science Fiction, Thriller, War, Western.

In this study, the goal of the movie ontology is to extract four different aspect-based contextual features – story, direction, cast performance, visual effects, and respective sentiment polarity for all movies. All 42 movie reviews related to fifteen different genres from the movie review datasets would be manually analyzed to generate a list of seed words for mapping movies to their respective concepts. The table below presents a partial list of identified seed words corresponding to all four movie contextual features mentioned above.

Contextual Features	Seed words
Story	story, concept, plot, sub-plot, screenplay, script, events, storyline, ending, climax, narrative, portrayal, dialog, storytelling, fairy-tale

Direction	adaptation, directing, direct, directed, directs, direction, film maker, directional, directional debut, directorial, film making, product
Cast Performance	acting, performance, character, actor, support cast, acted, played, acts, portray, debutant, villain, performed, lead, actress, artist, role, hero, heroine, starring, singing, cast performances
Visual-Effects	visually, animatronic, visual effects, animation, CGI, visual, graphics, animation, animate, animated, digital effect, graphic, photographed

Table 1. *List of seed words corresponding to all four aspect-based contextual features.*

The movie ontology can be constructed automatically from text using part-of-speech and dependency parsing. The extraction of entity pairs from grammatical patterns is fast and scalable to large amounts of text using NLP library `spaCy`. A specific function `get_entity_pairs(text)` is created to define entity pairs as entities/noun chunks with subject — object dependencies connected by a root verb. Other rules-of-thumb can be used to produce different types of connections. This kind of connection can be referred to as a subject-predicate-object triple. The derived movie ontology (Figure 2) is visualized by using the `NetworkX` library. The ontology is illustrated as a directed multigraph network with nodes sized in proportion to degree centrality.

Classification Methods

The Support Vector Machines is a classifier that finds best hyper plane between two classes of data, by separating positive and negative examples through the decision line, a solid line in the middle. Gap between the solid and dashed lines reflects the margin of movement of decision line left or right without miss-classification of document. On the other hand, Naïve Bayes classifier is basically a probabilistic classifier based on hypothesis. On the basis of assumption and training document; Bayesian learning is to find most appropriate assumption based on prior hypothesis and initial knowledge. Main assumption is that terms in test document have no relation among them and probability is calculated that document belong to a specific class. While the Naïve Bayes classification model is based on Bayes Theorem with the assumption of all variables as conditionally independent, the logistic regression model splits feature space linearly, and typically works reasonably well even when some of the variables are correlated (Ottesen, 2017). Last but not least, a bidirectional long short-term memory (Bi-LSTM) network is utilized for classifying the sentiment of each new movie review document. Due to use of gated functions, Bi-LSTM can effectively learn implicit knowledge from sequences by avoiding the troublesome of gradient vanishing. In addition, Bi-LSTM models enable additional training by traversing the input data twice. For performance measure, the confusion matrix is used for evaluating the prediction accuracy of each classification model (Figures 9~12). The classification accuracy metrics is calculated by actual labels that are equal to predicted label divided by total corpus size in test data.

3. Results

3.1 Overall Result

Support Vector Machine classification method classifies 28 out of 42 new movie review documents as negative reviews, whereas the rest of 14 movie review documents are categorized as positive reviews; the Logistic Regression method classifies 30 out of 42 new movie review documents as negative reviews, whereas the rest of 12 movie review documents are categorized as positive reviews. Meanwhile, the Naïve Bayes classification method classifies 29 out of 42 new movie review documents as negative reviews, whereas the rest of 13 movie review documents are categorized as positive reviews. Finally, the Bi-LSTM classification model classifies 8 out of 42 new movie review documents as negative reviews, whereas the rest of 34 movie review documents are categorized as positive reviews. The experimental result of the Bi-LSTM model seems pretty close to the manually labeled results which suggest that there are 13 movie review documents (Figure 13) that contain negative sentiments based on the sentiment polarity scores associated with all four different item-based contextual features – story, direction, cast performance, visual effects.

The experimental results (Table 2) have shown that Bi-LSTM classification model for movie reviews has achieved the highest classification accuracy (89%) in comparison with other classification methods. On the contrary, Naïve Bayes classification method has got the lowest average accuracy values (78%). The Support Vector Machine classification method for movie review data achieves the same F-measure score as the Logistic Regression method. However, its average accuracy score is 1% ~ 2% lower than that of the Logistic Regression method, although the difference is not statistically significant. Meanwhile, the Bi-LSTM classification model for movie review data achieves 10% ~ 11% higher average of classification accuracy than the Naïve Bayes model. The validation accuracy of the Bi-LSTM model increases as the number of epochs gets larger, but due to computing capacity restraint, the Bi-LSTM model is trained for at most 160 epochs.

3.2 Analysis and Interpretation

The comparison of Bi-LSTM, Support Vector Machines, Naïve Bayes, and Logistic Regression models for text classification indicates that the Bi-LSTM classification model outperforms the other models with the highest classification accuracy (89%). Using the labeled movie review dataset with named entities recognized for training the model significantly improves the aspect detection and sentiment classification capability of the Bi-LSTM model. Nevertheless, the inconsistency between the experiment results and the manually labeled results suggests that sentiment lexicon-based methods tend to put more weights on identifying negative words for predicting sentiment polarity regardless of contexts. For example, since the movie *Argo* is also a political thriller, the plots described in the movie review inevitably contains a lot of words that seem to indicate negative sentiments. Notwithstanding, in fact, these words do not represent any negative opinion of the movie critics who evaluates the film. In contrast, although the movie *The Imitation Game* does not have dark and gloomy plots, this movie's review reveals the movie critics' negative opinion of the film, as evidenced by the word "dull" that is used by the critics to describe the movie. Overall, however, the experimental accuracy (43.45% positive) indeed improved with the ontology incorporated compared to the prediction accuracy (100% positive) obtained without the ontology.

4. Conclusions

The experimental results imply that sentiment lexicon-based methods are domain specific and are typically based on bag-of-words models which ignore the semantic composition problem. Because ontology aims to provide knowledge about specific domains that are understandable by both developers and computers. The experiment is carried out effectively, and the result is good. Therefore, it is rational and effective to employ ontology to opinion mining.

For future works, the same experiments can be carried out on different datasets like Wikitology, Wikipedia, Word Net, Open Project Directory (OPD) etc. and other DNN-based NLP models, such as Bidirectional Encoder Representations from Transformers (BERT) based models, which are pretrained on large domain specific corpus incorporating knowledge graph and can learn high-level interactions among deep latent features. In-depth analysis and comparison can be holding on the diverse datasets and sentiment polarity scoring techniques.

Bibliography

Ottesen, C. (2017, October 24). Comparison between Naïve Bayes and Logistic Regression. Retrieved April 25, 2020, from https://dataespresso.com/en/2017/10/24/comparison-between-naive-bayes-and-logistic-regression/

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *arXiv preprint cs/0409058*.

Appendix

Classification Models	**Performance**			
	Precision	*Recall*	*F-measure*	*Accuracy*
Support Vector Machine	0.84	0.79	0.78	0.81
Naïve Bayes	0.83	0.77	0.76	0.78
Logistic Regression	0.84	0.79	0.78	0.82
Bi-LSTM Model	0.79	0.89	0.84	0.89

Table 2. *Comparison of performance of classification models measured by confusion matrices.*

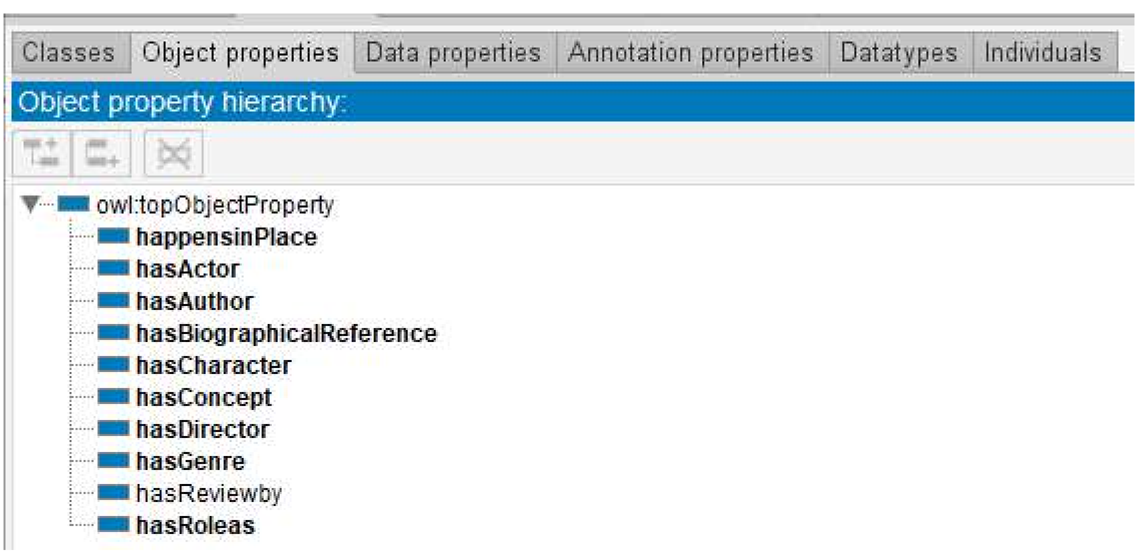


Figure 1. *OWL object property hierarchy of movie review ontology*

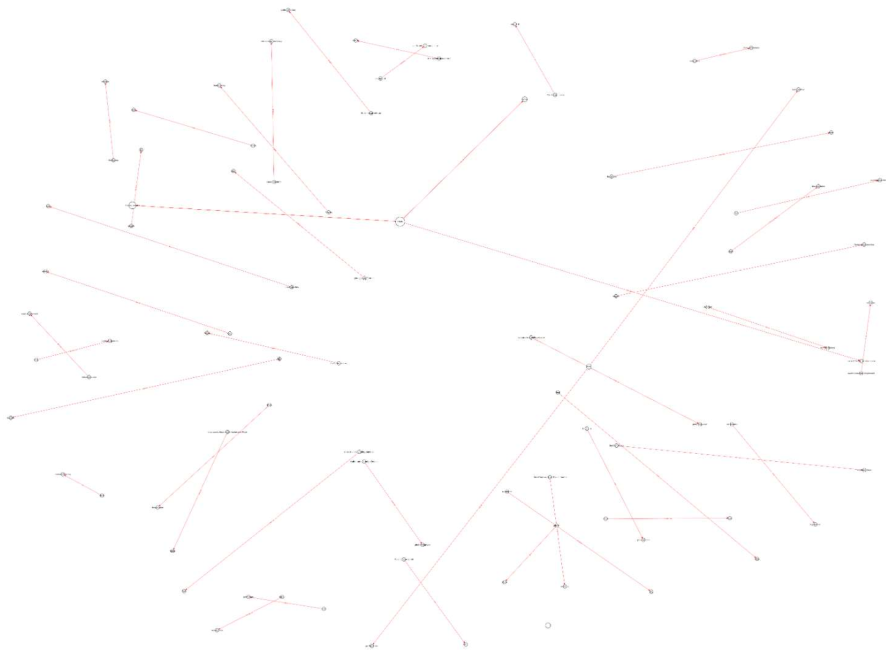


Figure 2. *The movie ontology generated from text using part-of-speech and dependency parsing is illustrated as a directed multigraph network with nodes sized in proportion to degree centrality.*

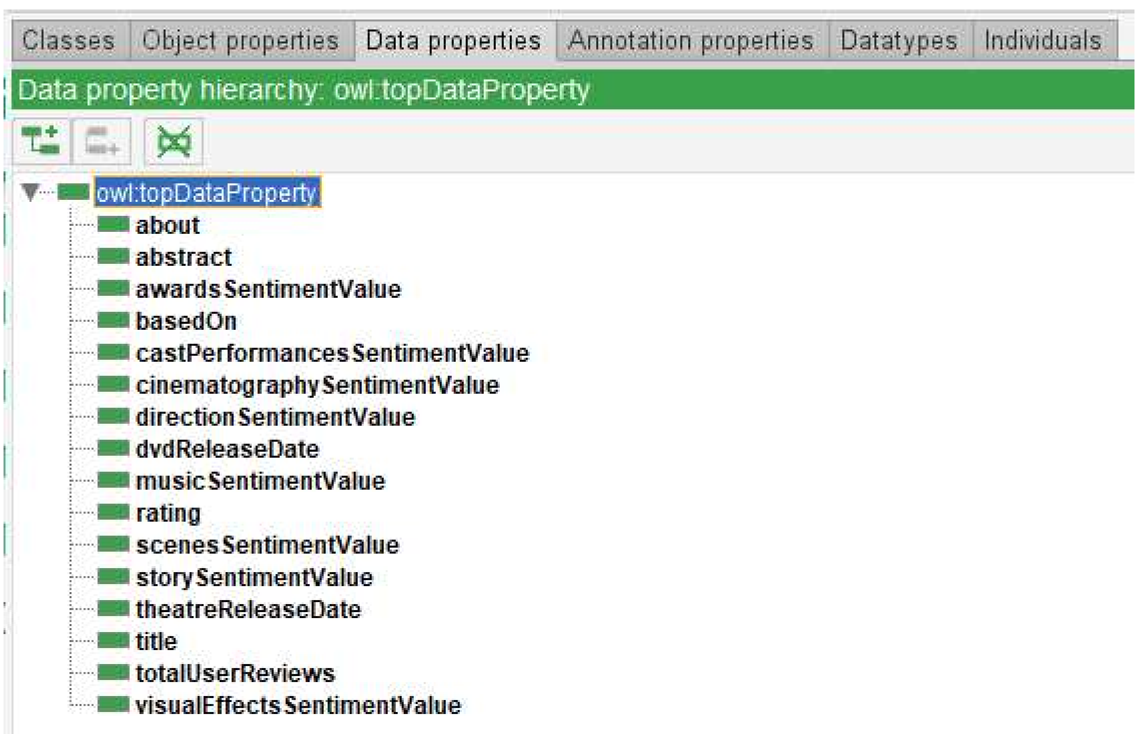


Figure 3. *OWL data property hierarchy of movie review ontology*

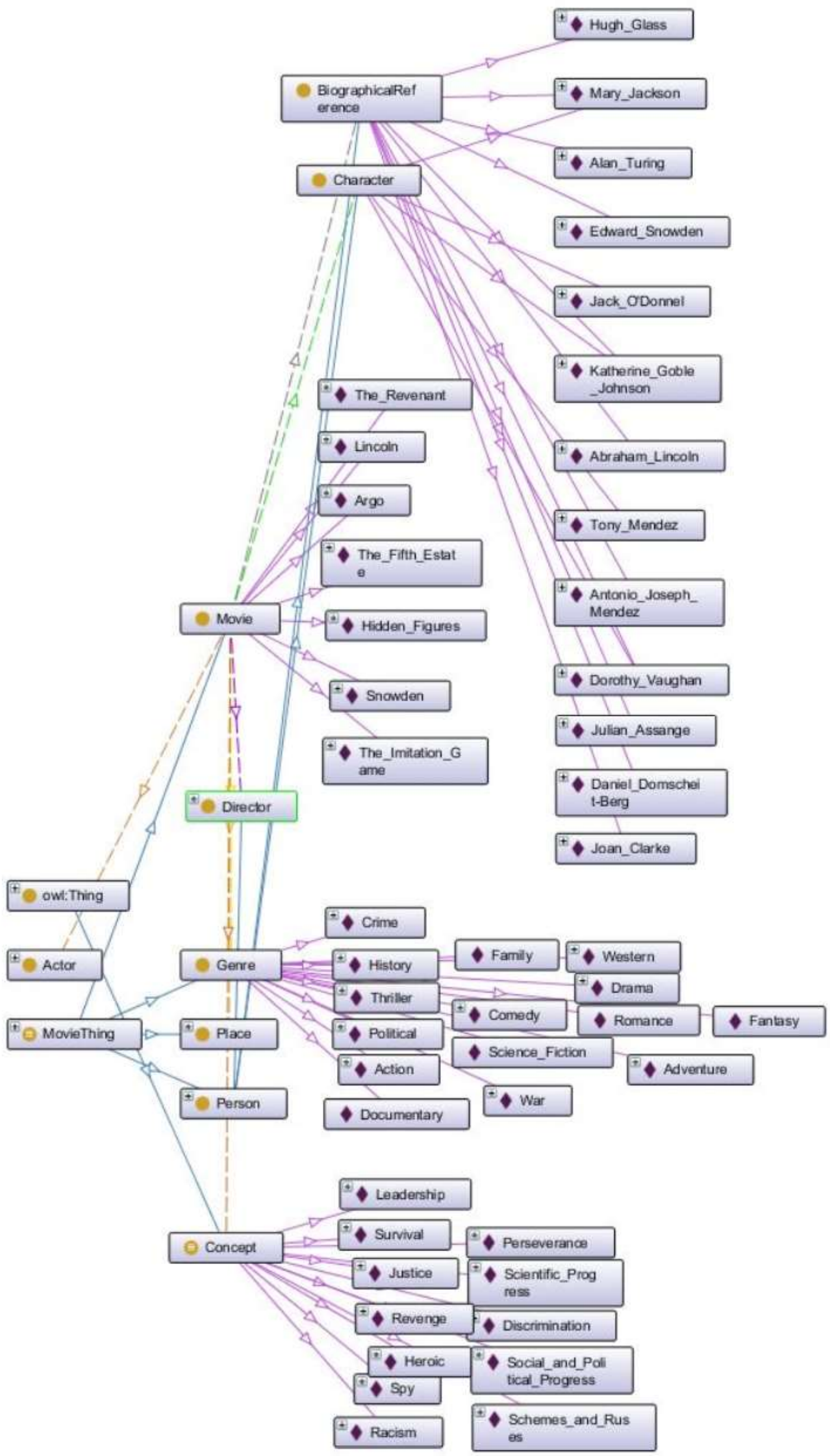


Figure 4. *OWL ontology of movie entities*

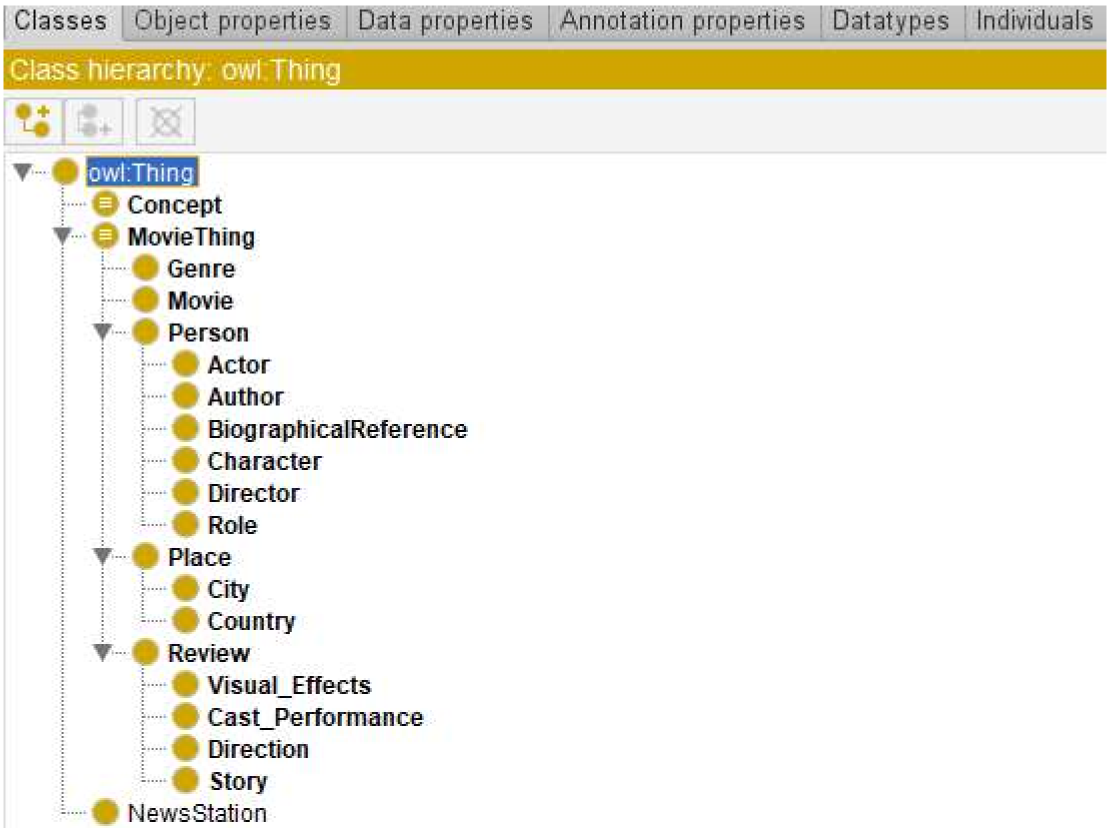


Figure 5. *OWL class hierarchy of movie review ontology*

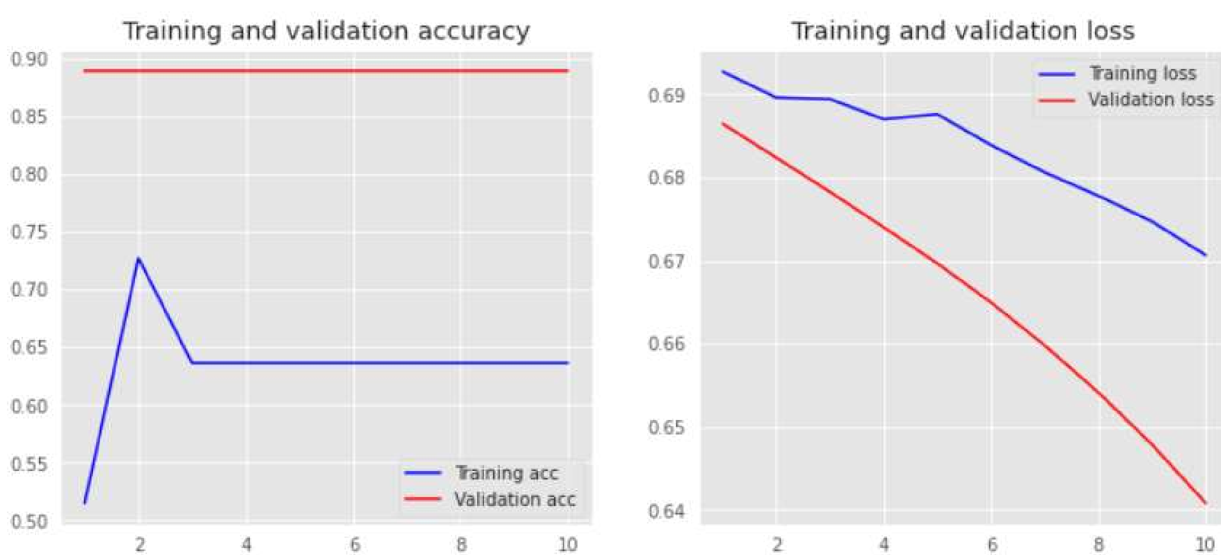


Figure 6. *Training and Validation Accuracy of Bi-LSTM Classification Model*

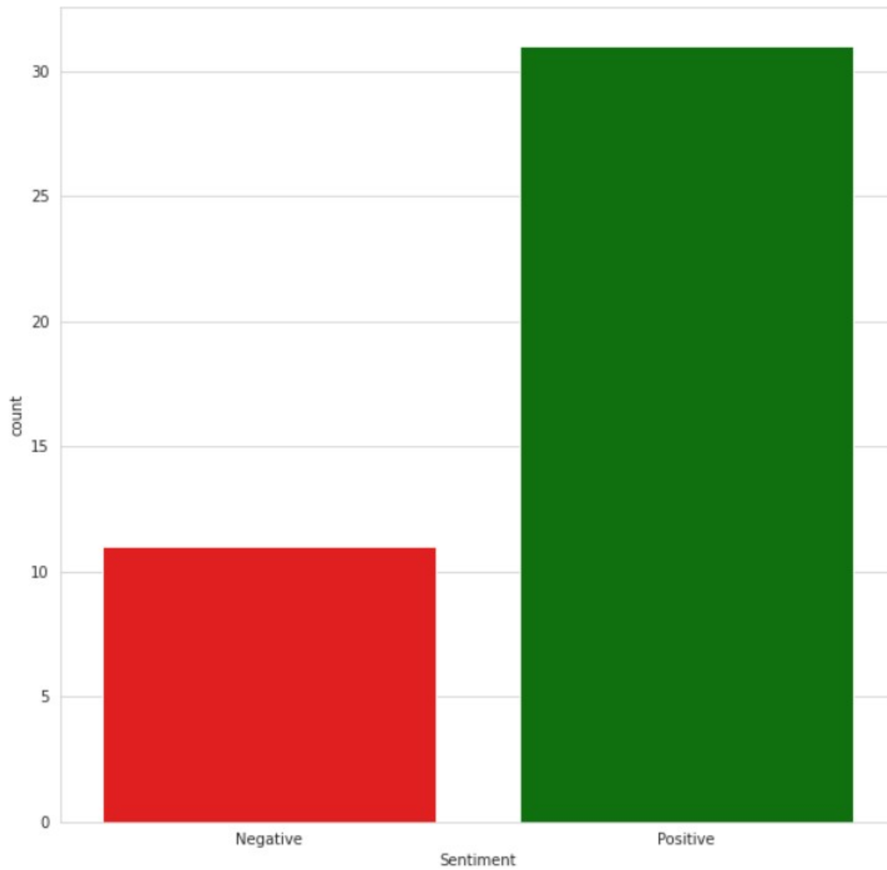
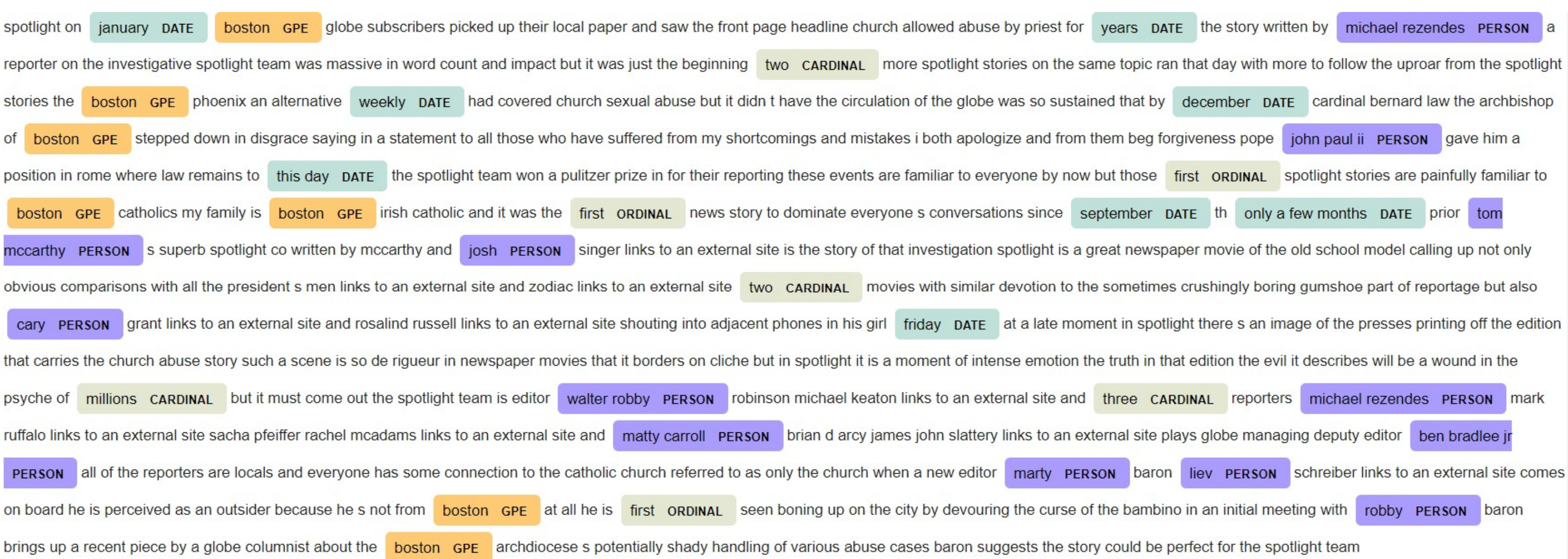


Figure 7. *Sentiment Analysis using VADER*



spotlight on january DATE boston GPE globe subscribers picked up their local paper and saw the front page headline church allowed abuse by priest for years DATE the story written by michael rezendes PERSON a reporter on the investigative spotlight team was massive in word count and impact but it was just the beginning two CARDINAL more spotlight stories on the same topic ran that day with more to follow the uproar from the spotlight stories the boston GPE phoenix an alternative weekly DATE had covered church sexual abuse but it didn t have the circulation of the globe was so sustained that by december DATE cardinal bernard law the archbishop of boston GPE stepped down in disgrace saying in a statement to all those who have suffered from my shortcomings and mistakes i both apologize and from them beg forgiveness pope john paul ii PERSON gave him a position in rome where law remains to this day DATE the spotlight team won a pulitzer prize in for their reporting these events are familiar to everyone by now but those first ORDINAL spotlight stories are painfully familiar to boston GPE catholics my family is boston GPE irish catholic and it was the first ORDINAL news story to dominate everyone s conversations since september DATE th only a few months DATE prior tom mccarthy PERSON s superb spotlight co written by mccarthy and josh PERSON singer links to an external site is the story of that investigation spotlight is a great newspaper movie of the old school model calling up not only obvious comparisons with all the president s men links to an external site and zodiac links to an external site two CARDINAL movies with similar devotion to the sometimes crushingly boring gumshoe part of reportage but also cary PERSON grant links to an external site and rosalind russell links to an external site shouting into adjacent phones in his girl friday DATE at a late moment in spotlight there s an image of the presses printing off the edition that carries the church abuse story such a scene is so de rigueur in newspaper movies that it borders on cliche but in spotlight it is a moment of intense emotion the truth in that edition the evil it describes will be a wound in the psyche of millions CARDINAL but it must come out the spotlight team is editor walter robby PERSON robinson michael keaton links to an external site and three CARDINAL reporters michael rezendes PERSON mark ruffalo links to an external site sacha pfeiffer rachel mcadams links to an external site and matty carroll PERSON brian d arcy james john slattery links to an external site plays globe managing deputy editor ben bradlee jr PERSON all of the reporters are locals and everyone has some connection to the catholic church referred to as only the church when a new editor marty PERSON baron liev PERSON schreiber links to an external site comes on board he is perceived as an outsider because he s not from boston GPE at all he is first ORDINAL seen boning up on the city by devouring the curse of the bambino in an initial meeting with robby PERSON baron brings up a recent piece by a globe columnist about the boston GPE archdiocese s potentially shady handling of various abuse cases baron suggests the story could be perfect for the spotlight team

Figure 8. *Visualization of named entities detected by spaCy in the "SD_Spotlight" movie review document*

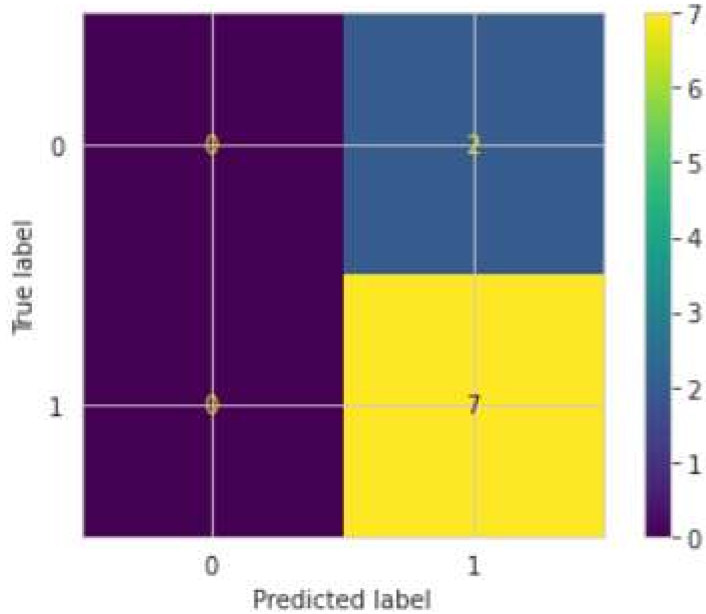


Figure 9. *Confusion matrix of the Support Vector Machine model.*

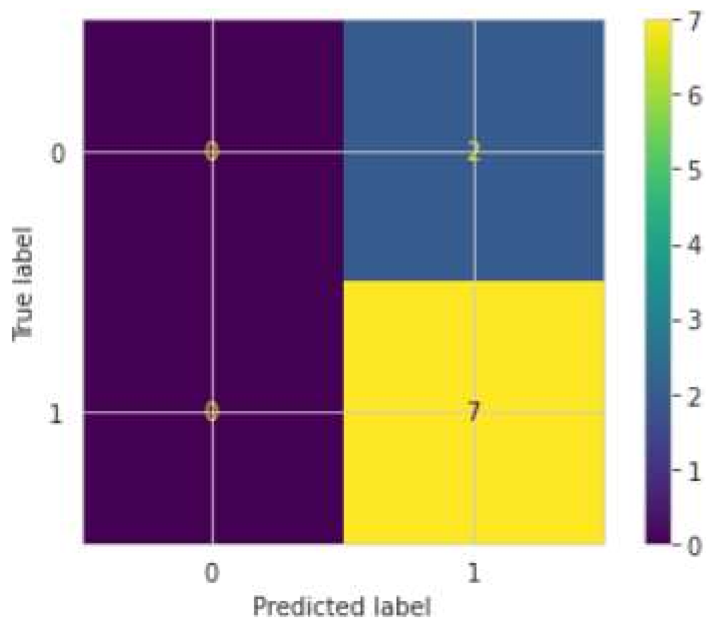


Figure 10. *Confusion matrix of the Logistic Regression model.*

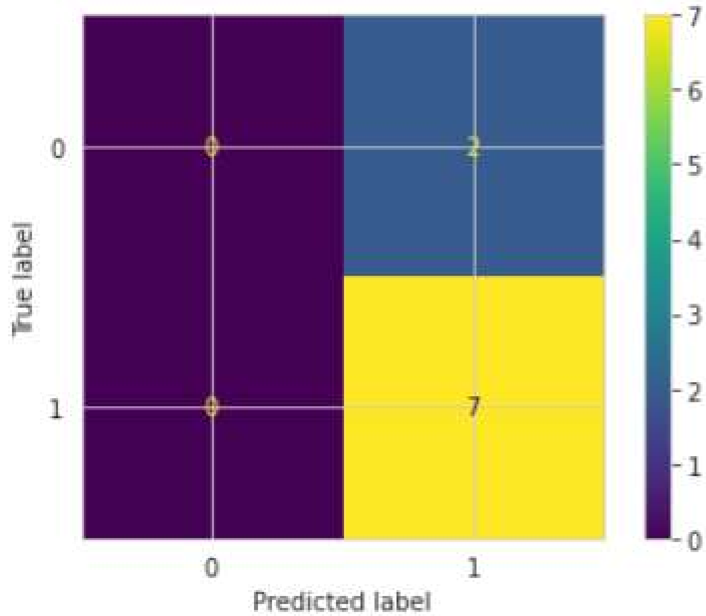


Figure 11. *Confusion matrix of the Naïve Bayes model.*

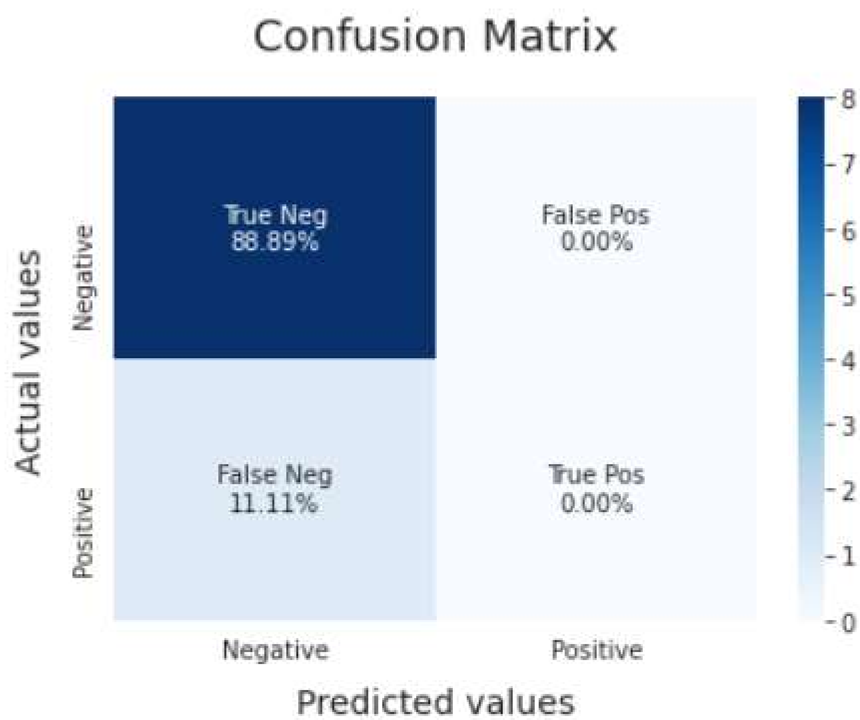


Figure 12. *Confusion matrix of the Bi-LSTM model.*

Doc_ID	DSI_Title	Text	processed	Polarity	Sentiment (VADER)	Manual Label	Aspects	svm	logit	naive_bayes	Bi-LSTM
0	KS_Doc1_:	In the	in the sprir	-0.2523	Negative	1	['spring', 'f	1	1	1	1
1	KS_Doc2_\	I am	i am writin	0.9405	Positive	0	['review', '	1	1	1	1
2	KS_Doc3_1	"The	the matrix	0.8928	Positive	0	['matrix', 'l	0	0	0	0
3	KS_Doc4_I	"I, Robot"	i robot tak	-0.9787	Negative	1	['robot', 'p	1	1	1	1
4	KS_Doc5_I	After a	after a rec	0.9974	Positive	0	['run', 'but	1	1	1	0
5	KS_Doc6_I	Spike	spike jonz	0.9957	Positive	0	['spike', 'jo	0	0	0	0
6	KS_Doc7_1	On Oct.	on oct a ne	0.965	Positive	0	['oct', 'new	1	1	1	0
7	PP_Doc1_	Title:	title disney	0.9973	Positive	1	['title', 'dis	0	1	0	0
8	PP_Doc2_(		title cinde	0.3973	Positive	0	['title', 'cin	0	0	0	0
9	PP_Doc3_,	Title:	title aladdi	0.9922	Positive	1	['title', 'ala	1	1	1	1
10	PP_Doc4_I		title reviev	0.996	Positive	0	['title', 'rev	1	1	1	0
11	PP_Doc5_I		title film r	0.9967	Positive	0	['title', 'film	1	1	1	0
12	PP_Doc6_1		title the ju	0.9983	Positive	0	['jungle', 'b	1	1	1	0
13	PP_Doc7_\		title what	0.9801	Positive	1	['title', 'girl	1	1	1	1
14	SW_DOC1	Title: AIR	title air for	0.9911	Positive	0	['title', 'air'	0	0	0	0
15	SW_DOC2	Title:	title clear a	0.9514	Positive	1	['title', 'dar	0	0	0	0
16	SW_DOC3	Title:	title crimsc	-0.9847	Negative	0	['title', 'crir	0	0	0	0
17	SW_DOC4	Title:	title hunt f	-0.8872	Negative	1	['title', 'hur	0	0	0	0
18	SW_DOC5	Title: IN	title in the	-0.9955	Negative	0	['title', 'line	1	1	1	0
19	SW_DOC6	Title:	title patrio	-0.8623	Negative	1	['title', 'pat	0	0	0	0
20	SW_DOC7	Title: THE	title the fu	0.9819	Positive	0	['review', '	0	0	0	0
21	VPD_DOC:	Paris: a cit	paris a city	0.9981	Positive	0	['city', 'ligh	1	1	1	0
22	VPD_DOC:	"I wanted	i wanted t	0.997	Positive	0	['aspect', 'l	1	1	1	0
23	VPD_DOC:	As beguilir	as beguilin	0.9985	Positive	0	['stroll', 'pa	1	1	1	0
24	VPD_DOC	Feel-good	feel good r	0.9975	Positive	0	['movie', 'i	0	0	0	0
25	VPD_DOC!	Existing so	existing so	0.9853	Positive	1	['saccharir	1	1	1	1
26	VPD_DOC(	"The Bucke	the bucke	-0.9553	Negative	1	['bucket', '	1	1	1	1
27	VPD_DOC	Frisky and	frisky and	0.9983	Positive	0	['woody', '	1	1	1	0
28	YF_Doc1_I	â€œHidd	hidden fig	0.9522	Positive	0	['figures', '	1	1	1	0
29	YF_Doc2_1	â€œThe	the imitati	0.9947	Positive	1	['imitation'	1	1	1	1
30	YF_Doc3_/	At one	at one poir	0.9903	Positive	0	['point', 'ar	1	1	1	0
31	YF_Doc4_S	Oliver	oliver ston	0.9855	Positive	1	['stone', 'p	0	0	0	0
32	YF_Doc5_1	The	the movie	-0.7182	Negative	0	['movie', 'r	1	1	1	0
33	YF_Doc6_I	After a	after a bru	-0.7351	Negative	0	['kinetic', 's	1	1	1	0
34	YF_Doc7_1	Mr.	mr assang	0.9657	Positive	1	['mr', 'assa	0	0	0	0
35	SD_12 yea	7) 12	years a sla	-0.8806	Negative	0	['years', 'sl	1	1	1	0
36	SD_Birdma	6)	birdman c	0.9979	Positive	0	['birdman'	1	1	1	0
37	SD_Green	2) Green	green boo	0.9977	Positive	0	['book', 'fo	1	1	1	0
38	SD_Moonl	4)	moonlight	0.9926	Positive	0	['moonligh	0	1	1	0
39	SD_Parasit	1)	parasite it	0.9983	Positive	0	['d', 'point'	1	1	1	0
40	SD_Shape	3) The	the shape	0.9708	Positive	0	['shape', 'v	1	1	1	0
41	SD_Spotlig	5)	spotlight c	-0.8945	Negative	0	['spotlight'	1	1	1	0

Figure 13. *Sentiment Labeling and Scoring of 42 new movie reviews.*