

# Sentiment Analysis of Amazon Canon Camera Review using Hybrid Method

Indrajeet Kaur Chhabra  
Department of Computer Engineering  
IET, Devi Ahilya University, Indore, India

Gend Lal Prajapati, PhD  
Department of Computer Engineering  
IET, Devi Ahilya University, Indore, India

## ABSTRACT

Sentiment Analysis or Opinion Mining is a new developing research field that has opened new challenges for researchers to be answered. Sentiment analysis or Opinion mining is a very important field in finding the correct sentiment of customer product review, election result analysis, summarization of news articles. Sentiment analysis has opened a new door in different domains like financial, telecommunication, business, medical, social events, and e-shopping. In this paper, a hybrid sentiment analysis approach is proposed to analyze “Amazon” Canon camera reviews and classify them into positive and negative polarity classes which is useful for other customers and organizations to take future decisions. The results of hybrid approach show improvement in accuracy, and also in precision and recall measures.

## Keywords

Sentiment Analysis, Opinion Mining, Feature Extraction, Machine Learning, Support Vector Machine, Amazon

## 1. INTRODUCTION

Internet has opened the new doors for people to buy products online and post reviews of the products that they purchase. With the increasing popularity of online shopping websites like “Amazon.com” the shopping experience changes in a way that the review regarding the products are available online and being generated by customers [2]. Before the origination of websites like “Amazon.com” for buying a product, people needs to decide whether to buy the product or not, they simply get the information through word of mouth, advertisements, family and friends. The organization needs to find opinion of people about products and services, for that they conduct the surveys and focused groups [14].

Human and machines are different in a way that people have an ability to understand personal opinions and the aim of artificial intelligence is to make the machine intelligent or make them think and take decisions. The field of computer linguistics that analyses opinions is called opinion mining or sentiment analysis (SA) [13, 17]. Sentiment Analysis also

known as opinion mining can be used to extract opinion from many kinds of texts, such as news, product reviews, social media blogs and messages [16]. In this paper a hybrid approach [12] of sentiment analysis is used, it includes the combination of lexicon-based approach and machine learning approach to classify the customer camera reviews more accurately.

There are different approaches to perform sentiment analysis as shown in Figure 1.

## 2. RELATED WORKS

### 2.1 Lexicon-Based Approach

The Lexicon based approach uses the built-in dictionaries of words and their semantic orientation to find the polarity of customer product reviews. Mukwazvure et al. [1] proposed a system that have three basic modules one is the crawling module to download the news and associated comments, second lexicon module(AFFIN-111) to identify the sentiments and last module is machine learning module (SVM classification and KNN classification) for training and testing the data, the data that is used to train was the data labelled by lexicon. Jose et al. [6] performs enhanced sentiment analysis on Twitter data to predict election result by using Word Sense Disambiguation, for that they have used three modules: Data acquisition module that is used to extract the twitter data, Pre-processing module to refine the tweets and include negation handling, sentiment classification module is used to classify twitter data using SentiWordNet and WordNet that is based on word sense disambiguation. Palanisamy et al. [7] proposed a Lexicon based Approach to sentiment analysis, they have created two types of a lexicon manually, common lexicon and category specific lexicon. Then pre-processing is performed including POS Tagging, Stemming, Exaggerated word shortening, Hashtag Detection and emotion detection. The pre-processed data is then used for sentiment calculation. R. Nithya et al. [10] suggested a sentiment analysis approach for unstructured review, utilizes a lexicon-based classifier SentiStrength to find word semantic orientation and strength.

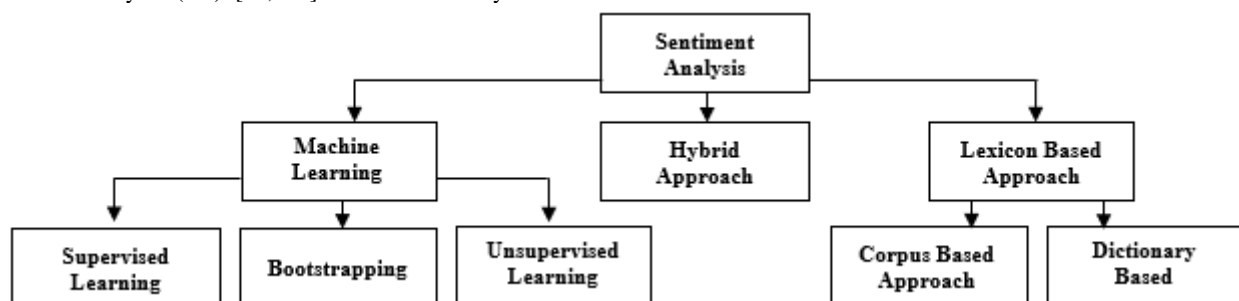


Figure 1. Approaches of Sentiment Analysis

### 2.2 Machine Learning Approach

The second technique used for sentiment analysis is machine learning that includes unsupervised and supervised machine

learning methods. Unsupervised learning methods do not require labelled data for classification while supervised learning algorithms require a labelled corpus for training the classifier. Fang et al. [16] proposed a sentiment analysis

approach for Amazon product review data by performing sentence extraction, POS tagging, identifying negation phrases and uses Naive Bayesian classifier, Support Vector Machine and Random Forest for sentiment classification. Choi et al. [2] developed a Domain-specific Sentiment Analysis technique using contextual feature generation they have employed a classification system consist of two parts: contextual feature generation and domain -specific sentiment classifier construction using bootstrapping approach. Khin phyu phyu shein et al. [3] they focus on feature level sentiment classification, having three main parts in their approach: assigning POS tagging, identifying domain related features by using domain Ontology and classifying domain related features by using linear SVM technique. To evaluate sentiment classification system, they have used precision and recall measures. Duncan et al. [4] employed Neural Networks to perform sentiment analysis on Twitter data. Pre-processing step is commonly done in all text classification techniques, in this paper they worked on feedforward neural network using MATLAB toolbox to train and test the tweets as positive or negative. Hassan A. Soliman et al. [5] mined online customer reviews by utilizing Support Vector Machine, this paper was introduced mainly to summarize the customer review. It works in four phases, Data Collection phase to get the review data from many websites like amazom.com, the collected data is then pre-processed to make it ready for classification in the next phase, then classification is performed on two different level, in level one features are classified and in level two polarity of sentence is classified. Santhosh Kumar K L et al. [11] aims to mine online reviews from websites like Amazon.com and classifies them into positive, negative and neutral using algorithms such as Naive Bayes classifier, Logistic Regression and SentiWordNet and compare the Recall, Precision and F-Measure values obtained for three different algorithms.

### 2.3 Hybrid Approach

The advances in sentiment analysis motivates researchers to explore the possibility of a hybrid approach that provides the

accuracy of machine learning based approach and speed of lexicon-based approach. Venugopalan et al. [8] explores sentiment analysis on twitter data by working on sequence of steps, first step is to retrieve data through Twitter API followed by data pre-processing steps that include Text Extraction, Spell-Checker, Slang replacement, and Link removal. They created labelled data to differentiate subjective and objective tweets, feature extraction is done by unigrams, then subjective and objective classification is performed. Finally, they proposed an approach for tweet classification by using Support Vector Machine and J48 classification techniques as they are widely used and give better performance than other techniques. Pujari et al. [9] performed comparison between various classifications techniques for feature-oriented sentiment analysis of product review data. First dataset of Amazon is downloaded, this downloaded data is pre-processed in next step, then POS tagging is used to identify words that depicts the sentiment of the review, in next step labelling of opinion words is performed using Bigram collection and at last polarity is determined using different classification techniques Naïve Bayes, Maximum Entropy Classifier and Support Vector Machine.

### 3. PROPOSED SCHEME

A large amount of opinionated data is available on various social networking sites, e-shopping websites, blogs and forums. Automated systems are required to search and classify opinions into positive and negative classes. In this proposed scheme, a hybrid method is used to perform the sentiment analysis [16, 20] on Canon camera reviews. In hybrid approach, first data pre-processing [13,18] is performed, then feature and opinion words are extracted, the extracted opinion words are then labelled as positive and negative with the help of Lexicon approach [15]. The labelled data is then used as training data for SVM classifier [1, 8, 15, 19]. This proposed scheme includes the main steps that must be performed for carrying out the sentiment classifications as shown in Figure 2.

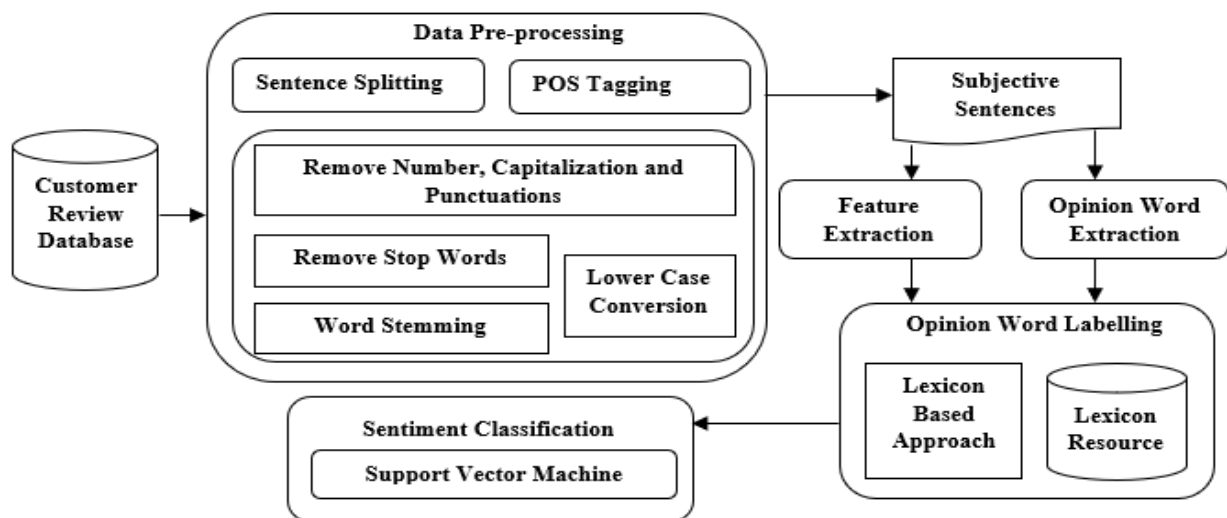


Figure 2. Proposed Scheme

## 4. EXPERIMENTAL SETUP

### 4.1 Data Collection

The reviews used in this work are collected from ‘Amazon.com’ website. The classification is performed on live review data. Canon camera reviews are collected as corpus.

### 4.2 Pre-processing of Data

To preprocess the data, Python’s Natural Language Toolkit have been used. NLTK is one of the important library used for natural language processing. First Python 3.5 32-bit version and NLTK library have been installed. The Toolkit has many different corpus, modules and packages to perform different

preprocessing task.

#### 4.2.3 Word and Sentence Tokenizer

NLTK has two methods `sent_tokenize`, `word_tokenize` to tokenize the reviews at sentence and word level. After this step the reviews are converted into tokens.

#### 4.2.2 Removal of Punctuations and Capitalization

Punctuations does not have any sentiments with them, so removed the punctuations from the reviews and also converted the text into lower case.

#### 4.2.3 Stop Word Filtering

NLTK includes an English stop word list, using this list the stop words were removed from the reviews, as stop words have no sentiments, so never useful in classification.

#### 4.2.4 Word Stemming

As words are in different inflected form, word stemming is used to convert the words into their root form. NLTK's Porter Stemmer is used to performed the word stemming.

#### 4.2.4 POS-Tagging

Part of speech tagging is one of the important step in review classification, it is used to assigned a part of speech tag to each word like verb, noun etc. NLTK's POS\_Tagger module is used to perform the part of speech tagging.

### 4.3 Lexicon Approach

Lexicon approach is a sentiment analysis approach that simply works on a lexicon dictionary, containing words along with their sentiment scores. In this work, SentiWordNet lexicon resource is used to find the positive and negative polarity of words. SentiWordNet lexicon gives three types of scores positive, negative and objective score. The total sum of all three score is 1.

### 4.4 SVM Classification

A hybrid approach is used in which the results of lexicon-based approach, the positive and negative polarity of the words is used to train the classifier. A support vector machine is used as it is a binary classifier so best worked for classification of reviews into positive and negative class.

The labelled data is spitted into two parts for training and testing. Different SVM have different kernels, the Linear SVM Kernel is used in this work.

## 5. RESULTS

A linear SVM classifier classifies the reviews into two polarity classes positive and negative. For classification it requires the training data to train the classifier and testing data to test the results. The labelled data is spliced into two parts, 30% to test and 70% to trained the classifier. For splitting the data, Python's `train_test_split` module is used. The classification results of hybrid approach have shown in Figure 3.

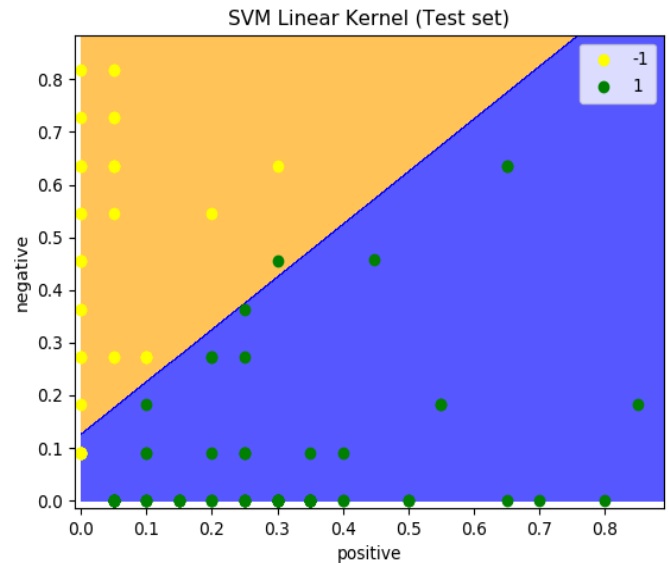


Figure 3. Plot of SVM Linear Classifier

### 5.1. Accuracy Measure

To calculate the accuracy, Python's `metrics.accuracy_score` module is used. The results shown in Table 1 proves that the accuracy of hybrid approach is better than the machine leaning approach(SVM).

Table 1. Accuracy Measure

No. of Reviews(Camera)	Technique	Accuracy (%)
500	SVM	72.72
500	Hybrid	94.31

### 5.1. Classification Report

To calculate the other measures like precision, recall, f1-score and support, the Python's `classification_report` module is used. The results of hybrid approach prove better as compared to SVM. as shown in Table 2.

Table 2. Classification Report

Technique	Class	Precision	Recall	F1-Score	Support
SVM	1	0.64	0.29	0.40	31
	-1	0.74	0.93	0.82	68
<b>Avg/Total</b>		0.71	0.73	0.69	99
Hybrid	1	1.00	0.84	0.91	25
	-1	0.92	1.00	0.96	49
<b>Avg/Total</b>		0.95	0.95	0.94	74

## 6. CONCLUSION

Sentiment analysis task is under research since the early 2000s and it is still in developing phase, especially the exploration of E-shopping websites, such as "Amazon.com". The experiments show that the proposed hybrid approach performs better than the machine leaning approach and lexicon-based approach. The future researchers have ample space for research in sentiment analysis field. Some unexplored areas are remained to explore to perform better in this field. They can use some other hybrid approaches to further improvise the accuracy of classification.

## 7. REFERENCES

- [1] Addlight Mukwazvure, K.P Supreethi, 2015. A Hybrid Approach to Sentiment Analysis of News Comments.2015 IEEE.
- [2] Yoonjung Choi, Youngho Kim, Sung-Hyon Myaeng, 2009, November. Domain-Specific Sentiment Analysis using Contextual Feature Generation.2009 ACM.
- [3] Khin Phyu Phyu Shein , Thi Thi Soe Nyunt 2010. Sentiment Classification based on Ontology and SVM Classifier. In 2010 Second International Conference on Communication Software and Networks, 2010 IEEE.
- [4] Brett Duncan and Yanqing Zhang,2015. Neural Networks for Sentiment Analysis on Twitter. In 14<sup>th</sup> International Conference on Cognitive Informatics & Cognitive Computing,2015 IEEE.
- [5] Taysir Hassan A. Soliman, Mostafa A. Elmasry et al., 2012, October. Utilizing Support Vector Machines in Mining Online Customer reviews. ICCTA 2012 IEEE.
- [6] Rincy Jose, Varghese S Chooralil, 2015, November. Prediction of Election Results by Enhanced Sentiment Analysis on Twitter Data using Word Sense Disambiguation. In International Conference on Control, Communication & Computing India (ICCC) 2015 IEEE.
- [7] Prabu Palanisamy, Vineet Yadav and Harsha Elchuri, 2013, June. Serendio: Simple and Practical Lexicon Based Approach to Sentiment Analysis. In Second Joint Conference on Lexical and Computational Semantics (\*SEM),2013 Association of Computer Linguistic (Vol. 2, pp. 543-548).
- [8] Manju Venugopalan, Deepa Gupta, 2015.Exploring Sentiment Analysis on Twitter Data. 2015 IEEE.
- [9] Chetana Pujari, Aishwarya and Nisha P. Shetty, 2018. Comparison of Classification Techniques for Feature Oriented sentiment analysis of product review data. In Data Engineering and Intelligent Computing, Advances in Intelligent Systems and Computing 542, Springer Nature Singapore Pte Ltd. 2018 Springer.
- [10] Mrs. R. Nithya, Dr. D. Maheshwari, 2014. Sentiment Analysis on Unstructured review. In International Conference on Intelligent Computing Applications,2014 IEEE.
- [11] Santhosh Kumar K L, Jayanti Desai, Jharna Majumdar, 2016. Opinion Mining and Sentiment Analysis on online customer review.2016 IEEE.
- [12] Orestes Appel, Francisco Chiclana, Jenny Carter and Hamido Fujita,2016. A Hybrid Approach to Sentiment Analysis. 2016 IEEE.
- [13] A.B. Pawar, M.A. Jawale and D.N. Kyatanavar, 2016. Fundamentals of Sentiment Analysis: Concepts and Methodology, Sentiment Analysis and Ontology Engineering, Studies in Computational Intelligence 639,2016 Springer.
- [14] Shirin Noekhah, Naomie Binti Salim and Nor Hawaniah Zakaria, 2018. A Comprehensive Study on Opinion Mining Features and Their Applications. In Recent Trends in Information and Communication Technology, 2018 Springer.
- [15] Bhavitha B K, Anisha P Rodrigues, Dr. Niranjan N Chiplunkar, 2017. Comparative Study of Machine Learning Techniques in Sentiment Analysis. In International Conference on Inventive Communication and Computational Technologies (ICICCT), 2017, IEEE.
- [16] Xing Fang and Justin Zhan,2015. Sentiment analysis using product review data. Journal of Big Data 2015 Springer.
- [17] Yoonjung Choi, Janyce Wiebe and Rada Mihalcea, 2017, June. Coarse-grained +/-Effect Word Sense Disambiguation for Implicit Sentiment Analysis. IEEE (2017)
- [18] A. Eesee and N. Omar, 2016, April. A Hybrid Method for Arabic Educational Sentiment Analysis. 2016 Journal of Applied Sciences.
- [19] Bissan Ghaddar, Joe Naoum-Sawaya, 2017, August. High Dimensional Data Classification and Feature Selection Using Support Vector Machine. 2017 European Journal of Operational Research.
- [20] Reinald Kim Amplayo, Min Song, 2017, March. An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews. 2017 Data & Knowledge Engineering.