# Chapter 19
# Knowledge Generation Using Sentiment Classification Involving Machine Learning on E-Commerce

**Swarup Kr Ghosh**

https://orcid.org/0000-0002-9312-4189
*Brainware University, Kolkata, India*

**Sowvik Dey**
*Brainware University, Kolkata, India*

**Anupam Ghosh**
*Netaji Subhash Engineering College, Kolkata, India*

## ABSTRACT

*Sentiment analysis manages the computational treatment of conclusion, notion, and content subjectivity. In this article, three sentiment classes such as positive, negative and neutral emotions have been demonstrated by appropriate features from raw unstructured data followed by data preprocessing steps. Applying best in class social analytics methodology to examine the sentiments embedded with purchaser remarks, encourages both producer and individual customers. Machine learning methods such as Naïve Bayes, maximum entropy classification, Deep Neural Networks were used upon the data, extracted from some websites such as Samsung and Apple for sentiment classification. In the online business arena, the application of sentiment classification explores a great opportunity. The subsidy of such an investigation is that associations can apply the proposed social examination framework to exploit the entire social information on the web and therefore improve their proper blueprint promoting strategies corresponding business.*

## 1. INTRODUCTION

Nowadays, large amount of unstructured data are accessible online from different E-commerce site, social network site, some forum such as movie review forum, travel blog, hotel blog etc. Native online users are in trouble to access the actual information as per their requirement since size of data are becoming vast day by day and more unnecessary information occupy space rather than significant information. So users are getting confused to handle these. In this work, authors have focused on different e-commerce site for a different product to the benefit of both customer as well as producer on the basis of user's review on a particular product. Thus sentiment classification takes place on the user's review followed by text mining. Hence, automated systems could be developed that could effectively organize and classify this data, so that it could be leveraged by human users in a meaningful way. Sentiment analysis in reviews is the way towards investigating a product review on the web to decide the general sentiment or on the other hand feeling about a product. Reviews speak for the supposed client produced substance, and this is of developing consideration and a rich asset for promoting groups, sociologists and analysts and other people who may be worried about feelings, sees, open temperament and general or individual attitude.

As such, sophisticated sentiment classification techniques that can automatically classify, on the basis of any form of data, for e.g. the analyzed travel blogs, whether the overall reviews of a specific destination either positive or negative would certainly be useful to users. Sentiment classification is a class of recently developed web mining techniques that can perform analysis on sentiment or opinions (Liu et al., 2005). Generally speaking, sentiment classification aims at mining text of written reviews from customers for certain products or services, and classifying the reviews into positive or negative or neutral opinions. The classification method has been used in the computing fields of information retrieval and natural language processing (Godbole et al., 2007). Again, there are challenges associated with mining data from texts (Go et al., 2009). In this domain specific area, word semantics in a particular review could contradict with the inclusive semantic direction (good or bad) of that review. For instance, if we take an example of travel blogs, an ''unpredictable'' camera implies a negative meaning to that camera; whereas a tour with an ''unpredictable'' experience is positive to explorers. Sentiment classification aims to extract the text of written reviews of customers for certain products or services by classifying the reviews into positive or negative opinions according to the polarity of the review (Dave et al., 2003). With the results of sentiment classification, consumers would know the necessary information to determine which products to purchase and sellers would know the response from their customers and the performances of their competitors. With the wide adoption of computing technology, sentiment classification of reviews has become one of the foci of recent research endeavors. The method has been attempted in different domains such as movie reviews, product reviews, customer feedback reviews, and legal blogs. Other potential applications include extracting opinions or reviews from discussion forums such as blogs, and integrating automatic review mining with search engines to automatically provide useful statistical data of search results or to build sentiment analysis systems for specific products or services.

The present research can be divided into two sections. The first section consists of pre-processing steps from raw data which are directly collected merchant or e-commerce websites on some product such as Nokia x6067, Canon-1200D-Digital Camera Black, Apple iPod and Samsung Galaxy Tab S8.4 LTE. Since this work is based on unstructured data, so data pre-processing takes a decisive part and which consists of some phases. The TF-IDF method has been applied to assign polarities on sentential data by using dictionaries of unigram, bigram and trigram for sentiment classification. The second section of the

## Related Content

A Review on Time Series Motif Discovery Techniques an Application to ECG Signal Classification: ECG Signal Classification Using Time Series Motif Discovery Techniques
Ramanujam Elangovanand Padmavathi S. (2019). *International Journal of Artificial Intelligence and Machine Learning (pp. 39-56).*
www.irma-international.org/article/a-review-on-time-series-motif-discovery-techniques-an-application-to-ecg-signal-classification/238127

Spatial Audio Coding and Machine Learning
Karim Dabbabi (2023). *Encyclopedia of Data Science and Machine Learning (pp. 2479-2498).*
www.irma-international.org/chapter/spatial-audio-coding-and-machine-learning/317688

MHLM Majority Voting Based Hybrid Learning Model for Multi-Document Summarization
Suneetha S.and Venugopal Reddy A. (2019). *International Journal of Artificial Intelligence and Machine Learning (pp. 67-81).*
www.irma-international.org/article/mhlm-majority-voting-based-hybrid-learning-model-for-multi-document-summarization/233890

The Digital Transformation of Rural Agricultural Business Management: A Case Study of China
Poshan Yu, Jiai Chen, Michael Sampatand Nyaribo Misuko (2022). *Empirical Research for Futuristic E-Commerce Systems: Foundations and Applications (pp. 23-52).*
www.irma-international.org/chapter/the-digital-transformation-of-rural-agricultural-business-management/309667

Genetic Data Analysis
M. Shamilaand Amit Kumar Tyagi (2021). *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning (pp. 358-372).*
www.irma-international.org/chapter/genetic-data-analysis/263327