A Survey on NLP Based Technique For Mining Product Reviews In E-Commerce Applications

Anita More[1], Priyanshu Jadon[2] , Dr. Durgesh Kumar Mishra[3]
[1]Dept of Computer Science & Engineering
[2] Professor, Dept of Computer Science & Engineering
[3] Professor & Director Dept of Computer Science & Engineering
[1, 2, 3] Sri Aurobindo Institute of Technology Indore, India

Abstract- *A significant roles of text mining and text processing is available in these days. That is not only suitable for developing scientific applications now in these days it also offering helping hands for the real world applications. The increasing data in different domains of information and technology increases the applicability of text mining and it's applications. In this paper the main aim is to include the relevant literature for data mining, text data mining and the sentiment based text analysis. In addition of that inclusion of the recent development in the domain of the sentiment based text analysis and relevant applications. Finally a data mining based sentiment analysis technique is proposed for discovering sentiments and emotions in the Amazon product review datasets. The proposed model in near future implemented and their performance is explored and presented using effective performance parameters.*

Keywords- sentiment analysis, NLP, natural language processing, survey, methodology design;

I. INTRODUCTION

Text mining played an essential role in processing and extracting data from unstructured source of information. The unstructured source of information is mostly involving the data in text format or text documents. Recently a new development in this domain is also observed in text data processing which is on the basis of hidden sentiments in text blocks such as social media post, reviews and others. These techniques help to understand the polarity of user for a particular topic, industry, brand or anything else. However the text mining is different from other kinds of text processing techniques such as text information retrieval and text data search. In this context when a text data is posted in some places the users are posting their sentiments with these data, the evaluation of the sentiments and emotions from the text data is also an essential part of text data analytics.

In this presented work the text data analysis is tried involve for preparing a supervised learning model for the evaluating the sentiments of reviews in an ecommerce platform. In this context the proposed work involves the understanding about the different concepts and methods which are useful for understanding the domain of text analysis and sentiment based text classification. The proposed work also includes the recent development and significant contributions in the domain of sentiment based text analysis using machine learning and data mining techniques. Further a data mining model introduced which accept the small text blocks and find the polarity of the users against a specific product or service offered by the targeted ecommerce platform. Here the sentiments are discovered in terms of positive, negative and neutral. Additionally it is proposed to work with the Amazon product review datasets. This section offers a basic overview of the proposed work involved in this paper. The next section introduces the detailed investigation of the target domain and relevant information.

II. BACKGROUND

This section provides the overview of the basic concepts and understanding about the sentiment analysis and other data mining concepts for finding the directions of the proposed research work.

A. Data Mining

The computerization has substantially enhanced our capabilities for collecting data from diverse sources. A tremendous amount of data has flooded every aspect of our lives. This growth in stored information has generated a dire requirement for new strategies and instruments that can keenly help us in changing the measures of information into data mining and learning. Data mining, is the advantageous extraction of examples speaking to information certainly put away or caught in huge databases, information distribution centers, the Web, other gigantic data storehouses, or information streams [1].

Data mining is a moderately new and promising innovation. It is the way toward utilizing the learning methods to naturally improve learning from information contained inside a database. The motivation behind data mining is to distinguish patterns and examples in data. It can be characterized as the way of finding significant relationship, examples, and patterns by diving into a lot of information using distribution center, machine learning, and artificial intelligence (AI), and perception methods. This technology should enable the discovery of trends and predictive patterns in data, the creation and testing of hypothesis, and visualizations [2].

B. Data Mining System

Data mining frameworks can be sorted by different criteria as [3]:

- **Data mining systems by the sort of data sources:** This is the kind of data managed, for instance, spatial data, sight and sound data, time-course of action data, content data, World Wide Web, etc.
- **Data mining systems demonstrated by the database used:** In perspective of the data included, for instance, social database, question arranged database, data appropriation focus, esteem based database, etc.
- **Data mining systems as the kind of pattern discovered:** In context of the sort of information found or data mining functionalities, i.e., depiction, division, alliance, portrayal, packing, etc. A system tends to be broad structures offering a couple of data mining functionalities together.
- **Data mining according to methodology used:** This is according to the data examination approach used, i.e., machine learning, neural frameworks, inherited counts, estimations, database arranged etc.
- **Other kinds of systems:** The classification can also take into account the degree of user involved in the data mining process i.e. query-driven systems, interactive exploratory or autonomous systems.

C. Challenges in Data Mining

This segment talks about difficulties in domain of data mining to handling and separating significant information that is examined [4] [5]:

- **Data Cleaning and Preprocessing:** It is a process to guarantee the data quality and to enhance the productivity and simplicity of the mining procedure, since true information have a tendency to be fragmented, conflicting, boisterous, high dimensional and multi-tactile and implies which are not appropriate for mining.
- **Post processing:** It is a procedure to refine and assess the learning from mining methodology. As illustration, it incorporates the disentanglement of the extricated learning, assessing the separated information and envisioning it. For the most part, Challenges in post-processing, (a) how to assess the examples (b) how to introduce the results to the area specialists in a way that is straightforward and (c) how to change over the examples into learning.

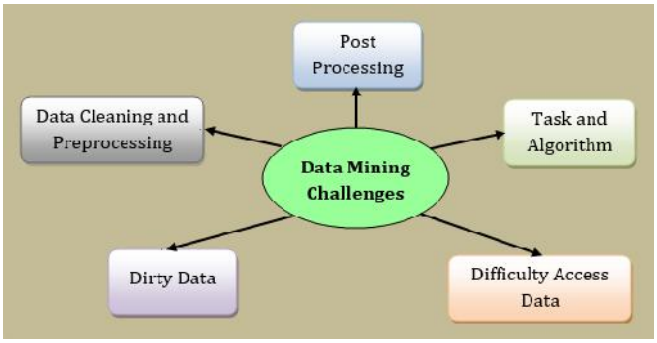

Figure 1 Data Mining Challenges

- **Task and Algorithms:** Data mining calculations are basic strides of learning. It normally includes an extensive variety of undertakings and calculations, i.e., design mining for finding intriguing affiliations and connections; bunching and drifts investigation to comprehend the nontrivial changes and patterns and characterization to counteract future. These diverse data mining task may utilize a similar database in various ways.
- **Trouble Accessing Data:** To getting to data that is normal and have a noteworthy test for data mining, for e.g. it is scattered throughout an association, more usually getting to information since it doesn't exist. Data mining by and large concurred that trouble getting to information because of the absence of an arrangement for information i.e. (1) how it can be obtained, (2) what information is needed, (3) how quality can be improved and (4) how it can be looked after. Data miners recommend working specifically with business clients to coordinate business issues with data prerequisites, and to utilize this as approach to start building up a more extensive arrangement for data gathering and availability.
- **Messy Data:** Here nothing unexpected that dirty information best the rundown, since it has been at the highest priority on the rundown for as far back as

quite a long while in region of data mining as a testing issue. Numerous information excavators gave contribution as: how they have attempted to conquer the issue and how to give a reasonable topic develops. Data miners utilize expressive insights and perception to help clients in understanding the data and distinguishing issue. Helping clients comprehend their information —hands on and causes everybody to pick up a mutual comprehension about the nature of information.

D. Text Mining

Text mining is a technique of data mining to capture meaningful information from unstructured format of data. That can be recognized as the process of analyzing text. It may be characterized as the process of analyzing text to extract information. Text is unstructured, and difficult to deal algorithmically. In modern culture, text is the most common vehicle for the exchange of information. The Text Mining function is the communication of factual information or opinions. The phrase "text mining" is generally used to denote a system that analyzes large quantities of natural text and detects lexical or linguistic patterns [6]. Data mining technology helps to extract data from various databases turned out to be numerical information, but unsuccessful when it came to text. Nowadays most of the information in business, industry, government and other institutions is stored in text form into database [7].

Text Mining is different from what we're familiar with in web search. In search, the user is looking for something that is already known. The problem is pushing aside all the material that currently isn't relevant to your needs to find the relevant information. In Text Mining, the goal is to discover unknown but useful information. It can be defined as a knowledge-intensive process in which a user interacts with a document collection by using a suite of analysis tools. In a manner to data mining, text mining seeks to extract useful information from data sources through the identification and exploration of patterns. In text mining, the data sources are document collections, and interesting patterns are found not among database records but in the textual data. Text mining appears to embrace the automatic natural language processing (NLP) and, arguably for example, analysis of linked structures such as citations in the academic literature and hyperlinks in the Web [8].

Text mining is an art of "text analytics" to make qualitative data usable by a computer. Qualitative data is descriptive and cannot be measured in numbers. Quantitative data is numerical and structured. For example, a photograph might be considered "qualitative" but when you break it down to the pixels, which can be measured. Text mining can help an organization derive potentially valuable business insights from text content such as documents, email and social media. Mining such data with NLP, statistical modeling and machine learning can be challenging, because natural text is often inconsistent. It contains ambiguities caused by syntax and semantics, language vertical and others [9]. It also referred to as text data mining, text analytics, or process of deriving high-quality information from text. Text mining process is as shown in following figure 2

Text mining process starts with a document collection from various resources. That would retrieve a particular document and pre-process it by checking format and character sets. Then it would go through an analysis phase. Analysis is semantic analysis to derive high quality information from text.

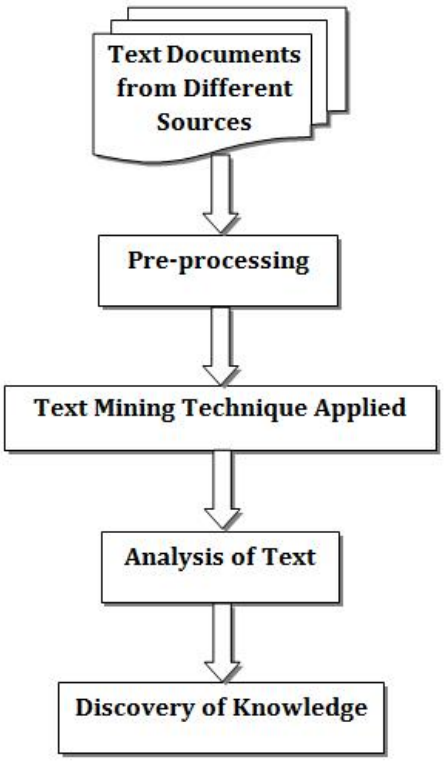

Figure 2 Text Mining Process

Many text analysis techniques are available; depending on goal of organization combinations of techniques could be used. Sometimes text analysis techniques are repeated for accuracy. The resulting information can be placed in an application system, yielding an abundant amount of knowledge [10, 11].

E. Social Media Text Analytics

The amount of digital information created and used is progressively rising with the growth of hardware and software. The real-world data come in a diversity of forms and can be bulky. This has augmented the need for powerful algorithms that can deduce and dig out facts and useful data. Text Mining has been successfully used for this purpose. It involves the process of structuring the input text (parsing, adding up some linguistic features and the removal of others, and successive insertion/deletions into the databases), deriving patterns within the data and end with interpretation and consideration of the output [12]. The word 'high-quality' in text mining refers to combination of originality, significance and interestingness. Text mining processes include document classification, clustering, ontology building, sentiment analysis, summarization, Information extraction etc. On the other side, information retrieval deals with crawling, parsing and indexing and retrieving documents [13]. An organization can use text analytics to expand insight into the content-specific values such as strength and significance of the data. It can be defined as the process of deriving information from text sources.

F. Part of Speech Tagging

A Part-Of-Speech (POS) Tagger is software that reads text in some language and assigns parts of speech to each word, i.e. noun, verb, adjective, etc. [14]. Automatic assignment of descriptors to the tokens is called Tagging. The descriptor is called tag. The tag indicates one of the POS, semantic information [15]. For Example:

- **Word:** Ram, ***Tag:*** Noun
- **Word:** Go, ***Tag:*** Verb
- **Word:** Good, ***Tag:*** Adjective

Note: some words can have more than one tags for example, chair can be noun or verb depending on the context.

POS tagger is a program that does this job. Taggers use several kinds of information: dictionaries, lexicons, rules, and so on. Dictionaries have category or categories of a particular word. That is a word may belong to more than one category. For example, run is both noun and verb. Taggers use probabilistic information to solve this ambiguity [15]. There are mainly two types of taggers: rule-based and stochastic. Rule-based taggers use hand-written rules to distinguish the ambiguity. Stochastic taggers are either HMM based, choosing the tag sequence which maximizes the product of word likelihood and tag sequence probability, or cue-based, using decision trees or maximum entropy models to combine probabilistic features. Ideally a typical tagger should be robust, efficient, accurate, tunable and reusable. In reality taggers either definitely identify the tag for the word or make the guess based on information. As the natural language is complex it is sometimes difficult for the taggers to make accurate decisions about tags. So that errors in tagging is not taken as a major roadblock to research.

G. Sentiment Analysis

Sentiment analysis offer organizations a fast and effective way to monitor the public feelings towards their brand, business, directors, etc. A wide range of features and methods for training classifiers for datasets have been researched in recent years. Sentiment analysis can be classified according to user emotions. Multiclass label sentiment analysis has been first introduced by Liu, B [16]. It is also known as opinion mining and subjectivity analysis. It is the process to determine the attitude or polarity of opinions or reviews written by humans to rate products or services. It can be applied on any textual form of opinions such as blogs, reviews and Micro-blogs. Micro-blogs are small text messages such as tweets [17]. Multiclass emotion analysis can be done on a document level or a sentence level is evaluated to determine the opinion polarity, where, the features describing the product/service. The second one, the document is divided into sentences each one is evaluated separately to determine the opinion polarity [18].

Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, and database sources through NLP. Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". The words opinion, sentiment, view and belief are used interchangeably but there are differences between them.

- ***Opinion:*** A conclusion to dispute (because different experts have different opinions)
- ***View:*** subjective opinion
- ***Belief:*** deliberate acceptance and intellectual assent
- ***Sentiment:*** opinion representing one’s feelings

Sentiment Analysis includes many tasks such as sentiment extraction, classification, and subjectivity classification, summarization of opinions, spam detection, and others. Mathematically, we can represent an opinion as a quintuple (o, f, so, h, t), where

o = object (An entity can be a, person, event, product, organization, or topic)
f = feature of the object o (An attribute of the object with respect to which evaluation is made);
so = orientation or polarity of the opinion (The orientation of an opinion on a feature f represent whether the opinion is positive, negative or neutral);
h = opinion holder (The holder of an opinion is the person or organization or an entity that expresses the opinion);
t = time when the opinion is expressed

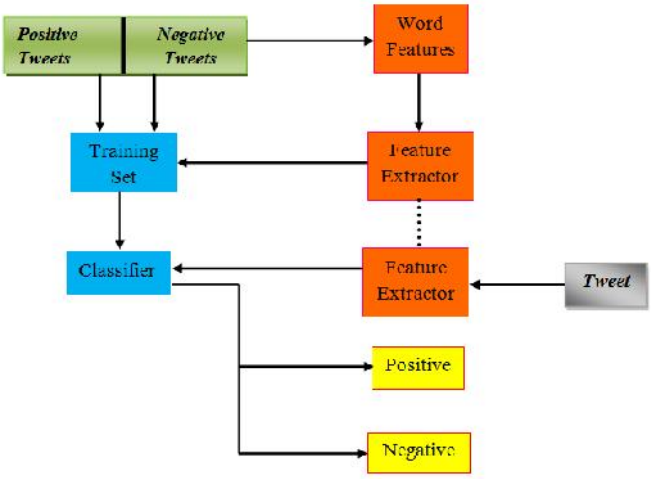

Figure 3: Sentiment Analysis Architecture

To train classifiers for sentiment analysis, we need to collect training data to apply appropriate learning algorithm with initial class Labels i.e. positive, negative, or neutral. An example model for sentiment analysis is given in figure 3 with the following phases for twitter data [18]:

A tweet contains a lot of opinions on data which are expressed in different ways by users. The twitter dataset labeled into two classes negative and positive. The raw data is having highly susceptible to inconsistency and redundancy. Pre-processing of tweet includes [19]:

- Remove all URLs, hash tags, and targets
- Correct the spellings; sequence of repeated characters is to be handled Replace all the emoticons with their sentiment.
- Remove all punctuations, symbols, and numbers
- Remove Stop Words
- Expand Acronyms
- Remove Non-English Tweets

The pre-processed dataset has many distinctive properties. In the feature extraction, we extract the aspects from the dataset. This aspect are used to compute the positive and negative polarity which is useful for determining the opinion of the individuals using models like unigram, bigram [20]. Machine learning techniques require representing the key features of text for processing. These features are considered as feature vectors which are used for the classification. Some examples features are:

- **Words Frequencies:** Unigrams, bigrams and n-gram models with their frequency are considered as features. There has been more research on using word presence rather than frequencies to better describe this feature.
- **POS Tags**: Parts of speech like adjectives, adverbs and some groups of verbs and nouns are good indicators of subjectivity. We can generate syntactic dependency by parsing or dependency trees.
- **Opinion Words and Phrases:** Apart from specific words, some phrases and idioms which convey sentiments can be used as features.
- **Position of Terms:** The position of a term with in a text can effect on how much the term makes difference in overall sentiment.
- **Negation:** Negation is an important but difficult feature. The presence of a negation usually changes the polarity of the opinion.
- **Syntax**: Syntactic patterns like collocations are used as features to learn subjectivity patterns by many of the researchers.

H. Classification

Data classification is the process of organizing data into categories/groups in a way that data objects of same group are more similar from different groups. Classification algorithm assigns each instance to a class such that classification error will be least. It is used to extract models that accurately define important data classes within the given dataset. The data from different domains includes valuable information and knowledge which is often hidden. Processing the data and retrieving meaningful information from it is a difficult task. Data Mining is a wonderful tool for handling this task [21].

The term classification could cover any context in which some decision is made on the basis of currently available information and in this process some formal method repeatedly making such judgments in new situations. We shall assume that the problem concerns the construction of a procedure that will be applied to a continuing sequence of cases, in which each new case must be assigned to one of a set of pre-defined classes on the basis of observed attributes. The

construction of a classification procedure from a set of data for which the true classes are known has also been termed pattern recognition, discrimination, or supervised learning [22].

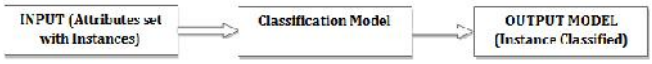

Figure 4 Classification Overview

Some of the urgent problems arising in science, and commerce can be regarded as classification or decision problems using very extensive data. Classification has been identified as a problem in the field of data mining, it try to find meaningful ways to interpret data sets. Classification technique, as predictive method is a supervised machine learning technique, and assuming the existence of a group of labeled instances for each object. The classification process is characterized by [22]:

- **Input,** a set of attributes with instances, including a class attribute, predictable;
- **Classifier** is used to predict the class of new instances;
- **Output**, a pattern classifier that classifies the instance in a certain category based on the other attributes

III. LITERATURE SURVEY

This section offers the recent developments and contributions of different authors in the domain of text data mining and natural language processing for sentiment based text classification.

X. Feng et al [23] aims at establishing a feature extraction model based on BiGRU and attention. Firstly, the vocabulary is vectorized by the skip-gram model. Then, to the pre-trained word vector, the sentiment words list can be reached and noise filtering would be conducted. Finally, the model extracts features using BiGRU and attentions. In this model, the attention layer is design to focus on the feature in different levels. This paper validates the model on two sentiment JD reviews and IMDB. Results show that the FGAttenBiGRU model achieves state of the art results on this task.

G. Vinodhini et al [24] compares neural network based sentiment classification methods (back propagation neural network, probabilistic neural network & homogeneous ensemble of PNN). They are validated using a dataset of product reviews collected from the Amazon website. An analysis is done to compare results of ANN based methods. The methods are evaluated using five quality measures and results show that the homogeneous ensemble of the NN method provides better performance. Among the two NN approaches used, probabilistic neural networks outperform in classifying the sentiment of the product reviews. The integration of NN based sentiment classification methods with PCA as a feature reduction technique provides superior performance in terms of time.

A. F. M. Agarap et al [25] attempts to understand the correlation of different variables in customer reviews on a women clothing e-commerce, and classify review whether it recommends the product and whether it consists of positive, negative, or neutral. To achieve these goals, author employed univariate and multivariate analyses on dataset features except for review titles and review texts, and they implemented a bidirectional RNN with long-short term memory unit for recommendation and sentiment classification. Results have shown that a recommendation is a strong indicator of a positive sentiment. On the other hand, ratings in product reviews are fuzzy indicators of sentiment. They found the bidirectional LSTM was able to reach an F1-score of 0.88 for recommendation, and 0.93 for sentiment.

T. Chen et al [26] propose a divide-and-conquer approach which first classifies sentences into types, then performs sentiment analysis separately on sentences. They find that sentences tend to be more complex if they contain more sentiment targets. Thus, they propose to first apply a NN based sequence model to classify sentences into three types according to the number of targets. Each group of sentences is then fed into a one-dimensional CNN for sentiment classification. The approach has been evaluated on four datasets and compared. Results show that: (1) sentence type classification can improve the performance of sentence-level analysis; (2) the approach achieves state-of-the-art results on benchmarking datasets.

The selection of right features needs to be carried out to select the best feature. ***A. Tripathy et al [27]***, the machine learning algorithm, i.e., SVM is used to select the best features from the data. These features are then given input to ANN, to process further. Different performance parameters i.e. precision, recall, f-measure, accuracy have been used on two different datasets.

Sentiment Analysis is a sub area of NLP which extracts user's opinion and classifies its polarity. ***V. Jha et al [28]*** are making attempt to create a sentiment aware dictionary. This dictionary is created using labeled data from the source domain and unlabeled data from source and target domains. This dictionary is used to classify the unlabeled reviews of the target domain. The work is carried out in Hindi.

When compared with labeling done by Hindi Sentiwordnet (HSWN), a general lexicon for word polarity, the method is able to label 23–24% number of words. The labels assigned by method and given by HSWN, for the available words, are compared and found matching with 76% accuracy.

Sentiment classification for reviews has attracted attention from the NLP community. By embedding prior knowledge into learning structures, classifiers often achieve a better performance. ***M. Wang et al [29]*** propose an algorithm based on deep learning and information geometry in which the distribution of training samples in the space is treated as prior knowledge and is encoded by DBNs. From the view of information geometry, authors construct the geodesic distance between the distributions of features for classification. The study contributes to the training of the DBN, since the distance is correlated to the error rate in the classification. Finally, evaluate proposal using data sets that are dedicated for sentiment classification. The results show that algorithm results in improvement over existing methods.

The most essential parts of sentiment analysis is the aggregation used to combine results from a lower level into an overall result. ***M. E. Basiri et al [30]***, the effects of the sentiment lexicon, aggregation level, and aggregation method on the sentiment polarity and rating classification of reviews are investigated. This aim, a new sentiment aggregation method based on the cross ratio operator is proposed. The results on four data sets show that the aggregation can improve classification.

The reviews posted by the consumers provide valuable information to other consumers. Such information is essential for decision making. This information is valuable not only for consumers to make decisions but also for businesses. ***A. S. Manek et al [31]***, a Gini Index based feature selection method with SVM classifier is proposed for sentiment classification for movie review data set. The results show that Gini Index method has better classification performance.

Aspect-based sentiment classification has shown promise in suppressing the noise. Not much research has been done on automatic aspect identification, and identification of implicit, infrequent and co-referential aspects, resulting in misclassifications. ***M. Afzaal et al [32]*** presents a framework of aspect-based sentiment classification that will identify the aspects very efficiently with high accuracy. The framework has been implemented as a mobile app that helps tourists, and performance has been evaluated by experiments on real-world datasets with excellent results.

IV. PROPOSED AIM

The proposed work is aimed to investigate and design an efficient and accurate model for sentiment based text analysis of review analysis. In this context a new model is proposed in this work for improving the existing methodologies using the extended feature computation and machine learning approach. The proposed model demonstrated in figure 5.

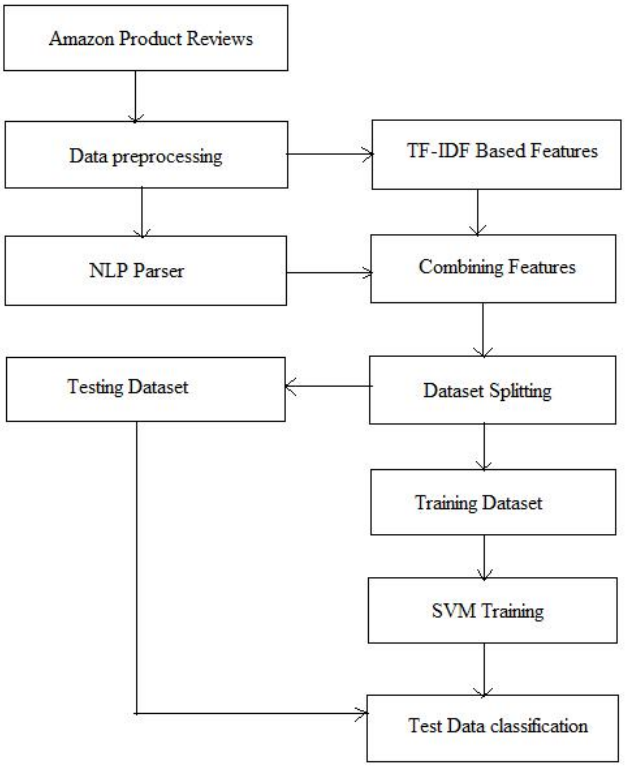

Figure 5 proposed model

The proposed model is demonstrate in figure 5, according to the given diagram the system accept a review dataset as input. This dataset is preprocessed in next phase for improving the data quality and enhancing the learning contents. Using this data the system extracts two different kinds of features first is based on NLP parser and second is the based on TF-IDF concept. Both kinds of feature vectors are combined in next phase. The combined features contain the NLP features as well as the essential keywords for processing the learning steps. The combined feature vector is further split into two parts first is used as the training data which contains 70% of samples and second part of data which contains 30% of samples are used as testing set. Using the extended features a popular machine learning model is trained namely SVM (support vector machine). After learning with the training samples the model is prepared for testing or final classification. In this context the test data is applied on the

system to get the product review orientation in terms of positive, negative or neutral.

V. CONCLUSION

In this paper the main objective is to explore the domain of sentiment analysis techniques and their applications. Additionally it is tried to find the effective methodologies to deal with the product reviews in an e-commerce platform. In this context a survey on recently contributed techniques and methodologies are carried out that offers new guidelines for achieving efficient and accurate technique for review analysis. During review of the existing techniques we observed that the text mining and NLP parser based techniques recently gain significant attention additionally effective for designing the sentiment based classification techniques. Finally using the obtained methods and algorithms a new technique is tried to design that is promising for classifying the user reviews according to their sentiments i.e. positive, negative and neutral. This methodology is implemented in future and their outcomes are demonstrated using the appropriate performance parameters.

REFERENCES

[1] J. Han, J. Pei, M. Kamber, *"Data mining: concepts and techniques"*, Elsevier, 2011.

[2] B. M. Ramageri, *"Data Mining Techniques and Applications"*, Indian Journal of Computer Science and Engineering, Volume 1 Number 4, pp. 301-305

[3] M. H. Dunham, *"Data mining introductory and advanced topics"*, Upper Saddle River, NJ: Pearson Education, New Delhi, 2003. Print. ISBN: 81-7758-785-4, 2006.

[4] I. H. Witten, E. Frank, M. A. Hall, "Data Mining: Practical Machine Learning Tools and Techniques", (3rd Ed.), Elsevier, 30 January 2011.

[5] Kantardzic, Mehmed, *"Preparing the data Data Mining: Concepts, Models, Methods, and Algorithms*", Second Edition (2003): 26-52.

[6] F. Sebastiani, *"Machine learning in automated text categorization"*, ACM Computing Surveys, Vol. 34, No. 1, pp. 1–47, 2002

[7] S. V. Gaikwad, A. Chaugule *"Text Mining Methods and Techniques"*, International Journal of Computer Applications, Volume 85 – No 17, January 2014.

[8] I. H. Witten, *"Text mining"*, 2004, available online at: https://www.cs.waikato.ac.nz/~ihw/papers/04-IHW-Textmini ng.pdf

[9] L. Kaushik, *"Text Mining - Scope and Applications"*, Journal of Computer Science and Applications, Volume 5, Number 2 (2013), pp. 51-55

[10] R. Talib, M. K. Hanif, *"Text Mining: Techniques, Applications and Issues"*, International Journal of Advanced Computer Science and Applications, Vol. 7 No. 11, 2016

[11] A. H. Tan, *"Text Mining: The State of the Art and the Challenges"*, in PAKDD99 Whorkshop on Knowledge Discovery from advanced Databases, Beijing, China, and April 1999.

[12] D. Zeng, et al. *"Social media analytics and intelligence"*, IEEE Intelligent Systems 25.6 (2010): 13-16.

[13] Nguyen, *"Text mining and Network Analysis of Digital Libraries in Data Mining Applications with R"*, (2014)

[14] Robin, *"Parts-of-Speech Tagging"*, available online at: http://language.worldofcomputing.net/pos-tagging/parts-of-speech-tagging.html

[15] S. M. Mohammad, S. Kiritchenko, and X. Zhu, *"NRC-canada: Building the state-of-the-art in sentiment analysis of tweets"*, in Proceedings of the seventh international workshop on Semantic Evaluation Exercises, 2013, pp. 321–327.

[16] V. A. Kharde, S. S. Sonawane, *"Sentiment Analysis of Twitter Data: A Survey of Techniques"*, International Journal of Computer Applications, Volume 139 – No.11, April 2016

[17] R. Socher, et al. *"Recursive deep models for semantic compositionality over a sentiment Treebank"*, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2013.

[18] R. Socher, et al. *"Recursive deep models for semantic compositionality over a sentiment Treebank"*, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2013

[19] B. Pang, L. Lee, *"A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts"*, 42nd Meeting of the Association for Computational Linguistics[C] (ACL-04). 2004, PP. 271-278.

[20] B. Pang, L. Lee, *"A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts"*, 42nd Meeting of the Association for Computational Linguistics[C] (ACL-04). 2004, pp. 271-278.

[21] Meenakshi, Geetika, *"Survey on Classification Methods using WEKA"*, International Journal of Computer Applications, Vol. 86, No.18, January 2014.

[22] S. Archana, Dr. K. Elangovan, *"Survey of Classification Techniques in Data Mining"*, International Journal of Computer Science and Mobile Applications, Volume 2 Issue 2, February 2014.

[23] X. Feng, X. Liu, *"Sentiment Classification of Reviews Based on BiGRU Neural Network and Fine-grained*

Attention", IOP Conf. Series: Journal of Physics: Conf. Series 1229 (2019) 012064

[24] G. Vinodhini, R. M. Chandrasekaran, *"A comparative performance evaluation of neural network based approach for sentiment classification of online reviews"*, King Saud University Journal of King Saud University – Computer and Information Sciences, 2016, 28, 2-12

[25] A. F. M. Agarap, P. M. Grafilon, *"Statistical Analysis on E-Commerce Reviews, with Sentiment Classification using Bidirectional Recurrent Neural Network"*, arXiv:1805.03687v1 [cs.CL] 8 May 2018

[26] T. Chen, R. Xu, Y. He, X. Wang, *"Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN"*, Expert Systems With Applications, 72, 2017, 221-230

[27] A. Tripathy, A. Anand, S. K. Rath, *"Document-level sentiment classification using hybrid machine learning approach"*, Knowl Inf Syst, Springer-Verlag London 2017, DOI 10.1007/s10115-017-1055-z

[28] V. Jha, Savitha R, P D. Shenoy, Venugopal K R, A. K. Sangaiah, *"A novel sentiment aware dictionary for multi-domain sentiment classification"*, Computers and Electrical Engineering, 2017, 1-13

[29] M. Wang, Z. H. Ning, C. Xiao, T. Li, *"Sentiment Classification Based on Information Geometry and Deep Belief Networks"*, 2169-3536, volume 6, 2018 IEEE

[30] M. E. Basiri, A. Kabiri, M. Abdar, W. K. Mashwani, N. Y. Yen, J. C. Hung, *"The effect of aggregation methods on sentiment classification in Persian reviews"*, Enterprise Information Systems, © 2019 Informa UK Limited, trading as Taylor & Francis Group

[31] A. S. Manek, P D. Shenoy, M C. Mohan, Venugopal K R, *"Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier"*, World Wide Web, © Springer Science+Business Media New York 2016

[32] M. Afzaal, M. Usman, A. Fong , *"Tourism Mobile App With Aspect-Based Sentiment Classification Framework for Tourist Reviews"*, IEEE Transactions on Consumer Electronics, Vol. 65, No. 2, May 2019