International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5914
Article Received: 12 October 2022 Revised: 18 November 2022 Accepted: 26 December 2022

Feature-Based Opinion Classification Using the KPCA Technique: Concept and Performance Evaluation

Sandeep Kumar and Bindiya Ahuja
Department of Computer Science and Engineering
Lingaya's Vidyapeeth
Faridabad, India
19phcs03w@lingayasvidyapeeth.edu.in

Abstract

Over the last several years, a widespread trend on the internet has been the proliferation of online evaluations written by people with whom they share their ideas, interests, experiences, and opinions. Opinion mining, also known as sentiment analysis, is the process of classifying pieces of text written in a natural language on a subject into positive, negative, or neutral categories according to the human emotions, views, and feelings that are communicated in that text. The field of sentiment analysis has progressed to the point that it can now analyse internet evaluations and provide significant information to people as well as corporations, which may assist these parties in the decision-making process. In the proposed model, feature extraction extracts the collection of features that are both semantically and statistically significant using the kernel principal component analysis (KPCA) method. According to the findings of the simulations, the suggested model performs better than other existing models.

Keywords: Opinion Mining, Feature-based Opinion, Kernel principal component analysis, CNN, Sentimental Analysis, LSTM, Feature Extraction.

I. Introduction

An increasing number of people are turning to the internet to voice their responses or opinions in the form of reviews, comments, or responses to questions posed on online forums, blogs, and social networking sites because of the rapid advancement of technology. This has led to an increase in the amount of user-generated content that is available on the internet [1]. The lightning-fast pace at which Web 2.0 has come into existence has been a driving force behind the meteoric rise in the number of websites devoted to online shopping and social media content. It has evolved into a venue where individuals may interact with one another, do commerce, provide comments or recommendations, and do a variety of other activities. Opinions and sentiment analysis are fundamentally an individual's attitude, feelings, or sentiments regarding a certain thing [2]. Sentence-level mining, document-level mining, and feature-level mining are all instances of how opinion mining can be performed. Sentence-level opinion mining works by first determining if a sentence is good, negative, or neutral, and then classifying each sentence accordingly. With document-level opinion mining, one considers that a single opinion holder is responsible for the document's content after determining the document's overall polarity [3]. Once the characteristics have been found, the views that match each feature are summarised in the feature-based opinion-mining process. It reveals which aspect of the product is most valued by buyers and which is least. Online retailers have a wide range of goods, each of which can be evaluated on its characteristics due to the plethora of consumer reviews written about it [4]. The mining of opinions is an effective model for staying focused on several business trends relating to the administration of transactions, management of status, and advertising respectively. The feedback from clients is also used when attempting to make predictions about patterns [5].

Opinion mining and sentiment analysis are subfields of natural language processing (NLP), information retrieval (IR), and text mining (TM) even though they all deal with the interpretation of the text and the expression of user sentiment. Sentiment analysis is the automated analysis of opinionated material to determine its polarity, whereas opinion mining is the process of collecting people's views and ideas expressed about entities or features/aspects of entities from unstructured texts [6].

1.1 Feature-based opinion mining

Opinion mining based on product features involves learning which characteristics customers value most and which they find most disagreeable. Figure 1 illustrates the three possible text-mining techniques for sentiment analysis that could be utilized in feature-based opinion mining such as a word-based approach, a pattern-based approach, and an ontology-based approach. It also includes an exploration of potential combination strategies that might boost outcomes with less work [7][8].

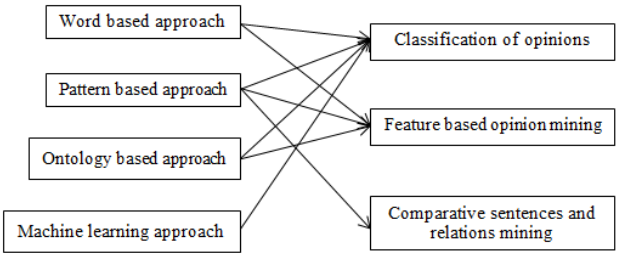
Word based approach
Pattern based approach
Ontology based approach
Machine learning approach
Classification of opinions
Feature based opinion mining
Comparative sentences and relations mining

Figure1. Various methods utilized in specific opinion mining analysis.

- **Word based approach**

Word clouds, also known as tag clouds, are a visual demonstration of text data that are often used to display keyword information (tags) on websites or to visualize free-form text. Typically, a tag would be a single word, and the relative significance of that tag is shown by its size or color. The sentiment analysis based on the cloud of tags can be performed using the feature. The items in the "Cons" cloud are assumed to be drawbacks, whereas those in the "Pros" cloud are taken to be positive aspects of the product. First, the stop list has to be cleared out so that the cloud of tags may be produced. Those expressions have the potential to cloud the outcomes of this strategy [9].

- **Ontology-based approach**

Ontology is a standardized and agreed-upon definition of a topic of study. Knowledge is represented formally as a collection of ideas and their connections within a domain. Ontological classes (notions) can be organized in a tree-like hierarchy. A collection of objects (individuals, instances of ideas) that stand in for actual things or beings in a particular area may also be found in the ontology. Attributes elucidating the characteristics of the notions are possible. In addition to describing the domain, ontology could be used to reason about the items found inside it. The ontology, built in the form of a graph, facilitates opinion mining that is based on a set of features. Sentiment analysis may benefit from ontology in several ways. In the standard model, a single viewpoint may serve as an ontology example [10].

- **Pattern-based approach**

A pattern-based approach makes it possible to recognize certain terms from people's perspectives. Some sentiments could be attached to such sentences. One of the advantages of using this method is that it makes it possible to identify sentences that affect the sentiment in many ways, such as by negating, nullifying, strengthening, and so on. Patterns are used in feature-based sentiment analysis to help discover the qualities of a product and combine those attributes with phrases that have a positive or negative valence.

The remaining part of the research is laid out as follows. A review of research related to the study is presented in section 2. Section 3 and 4 demonstrate the background study and problem formulation, whereas section 5 & 6 illustrate the research methodology for feature-based opinion using the KPCA algorithm and also determines the techniques, dataset, and proposed framework. Sections 7 &8 reveal the results, and conclusion.

II. Review of literature

In this part, various researchers present their results and techniques, expanding on the prior study of a design and implementation of a performance evaluation of feature-based opinion using KPCA techniques.

Srivastava and Vijay et al., (2022) [11] observed that sentiment analysis has been one of the most important areas of academic research since the proliferation of social media platforms on the internet. There is a lot of fervour in human existence as a direct result of the fact that millions of people share their thoughts, ideas, emotions, feelings, and views on social media websites such as Twitter, YouTube, and Facebook, amongst others. Sentiment analysis focuses largely on the categorization and forecasting of the views of individuals concerning a certain issue which is sometimes known as opinion mining. It comprises placing written documents or words into categories that are either good or negative depending on the viewpoint that is given on a certain subject. Even though sentiment analysis seems to be equivalent to text classification, it faces some challenges that have generated a significant amount of study in this field. Various ML and lexicon-based algorithms have been built during the development of the story to automate the work of sentiment analysis. This work provides the outcomes of a tertiary study that investigates the degree of research that is currently being conducted. This information must be preserved to enable future researchers to develop new automated methods that are capable of solving all of the issues and producing the best possible outcomes.

Kumar et al., (2022) [12] studied that sentiment analysis/opinion mining is a method for analysing how people feel about various things, such as brands, companies, services, problems, people, and events. It's a vast area with significant drawbacks. People have a propensity for sharing their thoughts about anything from a particular item or service to a general subject or even a specific person, group,

DOI: https://doi.org/10.17762/ijritcc.v10i2s.5914
Article Received: 12 October 2022 Revised: 18 November 2022 Accepted: 26 December 2022

or occurrence. Data mining, tasks including image classification, image retrieval, cluster analysis, and pattern recognition all make use of the idea of feature selection. This method not only improves the precision of the data analysis processes but also reduces the computing costs that are spent in the process. Words, phrases, signs, and symbols are all considered signifiers, and the focus of the Semantic characteristics should be on the connections between them. Linguistic semantics is a subfield of semantics dedicated to the study of human-based expressions, such as that seen in online forums and personal blogs. The presented semantic-based techniques assist reduce the number of characteristics needed to examine each word's prediction capacity and the attributes used to choose such words. Experiments were run using the Naive Bayes, FLR, and AdaBoost classifiers, and the results were associated to assess the feature selection strategies.

Pradeepa et al., (2021) [13] intended that users' feature-based online comments have become more important to both customers making purchases and companies searching for feedback on their products on e-commerce platforms. Therefore, it's vital to develop opinion-mining platforms centred on eliciting customers' feature-based assessments of items. E-commerce service providers confront the challenge of sifting through massive volumes of data to identify customer sentiment. An improved global Hypergraph for opinion mining using the Helly property method and the Hadoop distributed computing framework. State-of-the-art methods are used to evaluate performance, and the results show that the proposed feature-based opinion mining system is superior to the alternatives in terms of both accuracy and time complexity. Experimental findings demonstrate the efficacy of the proposed method for extracting aspect-sentiment, categorizing, and summarising evaluations of products posted on the web.

Alamoudi et al., (2021) [14] examined that information gathering and making sound choices have been greatly aided by opinion mining. It has analysed both the language of reviews and their rankings to determine what makes for useful information for consumers. Binary (positive and negative) and ternary (somewhere in between) classifications of Yelp restaurant reviews have been conducted (positive, negative, and neutral). Predictive models using ML, deep learning, and transfer learning have all been used. Furthermore, it suggests a novel unsupervised strategy to deploy for aspect-level classification tasks based on similarity measures to make use of the powerful potential of pre-trained language models like GloVe and to avoid many of the challenges associated with the supervised learning models. A final accuracy of 98.30% was achieved with the ALBERT model. The suggested approach for extracting aspects was successful to the tune of 83.04% of the time.

Ahamed et al., (2019) [15] evaluated that the emergence of social media has provided a cannon for the proliferation of digital data. In sentiment analysis, enormous amounts of textual data, such as those seen in internet reviews, are evaluated to determine the underlying emotions at play. It offered a method for doing sentiment analysis on web data that makes use of a fuzzy-based ML algorithm to perform nuanced analysis on voluminous online views by the impact of linguistic hedges. SentiWordNet was used to create the seven dimensions which are composed of a pre-processing step, a feature-selection stage, and a fuzzy-based sentiment analysis stage. The categorization process makes use of some machine learning techniques, including Naive Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbor. Jsoup is used to collect user feedback from throughout the web, which is then subjected to a steaming and tagging procedure. This fuzzy-based technique is studied for the Mobile and Laptops dataset, and it is also evaluated with state-of-the-art techniques, which offer an upper suggestion of 94.37% accuracy and reliability using Kappa indications displaying fewer error rates. The results of the inquiry are validated on the data used for training using a ten-fold cross-validation method, which accomplishes that this method may be successfully used for sentiment analysis and utilized as an aid for making decisions online.

Zainuddin et al., (2018) [16] evaluated that websites and programs dedicated to social media such as Facebook, YouTube, Twitter, and blogs have emerged as some of the most popular forms of online entertainment. As a result of the vast amounts of information that can be gleaned from this medium, it has emerged as an appealing resource for associations that want to keep tabs on the viewpoints held by users, and as a result, it is garnering a great deal of attention in the area of sentiment analysis. In the early days of sentiment analysis, researchers focused on determining the overarching tone of a text rather than parsing out the specifics of a sentiment's expression within it. This study performed a more fine-grained analysis by considering the usage of aspect-based sentiment analysis on Twitter. This allowed for a more accurate representation of the data. It is suggested to include a feature selection approach to create a new hybrid sentiment categorization for Twitter. It has evaluated the features selection strategies of PCA, LSA, and Random Projection (RP) and compares their classification accuracy. The assessment using other classification algorithms also showed that the novel hybrid technique gave

DOI: https://doi.org/10.17762/ijritcc.v10i2s.5914
Article Received: 12 October 2022 Revised: 18 November 2022 Accepted: 26 December 2022

relevant findings, and Twitter datasets were used to evaluate the hybrid sentimental analysis, given the varied domains represented by the datasets. The results of the implemented systems demonstrated that the novel hybrid sentiment classification improved accuracy performance by 76.55, 71.62, and 74.24% over the previous baseline sentiment classification techniques.

Babu et al., (2017) [17] analysed that e-commerce relies heavily on social media analytics to glean information that is relevant to a product or service to make informed purchasing decisions. Mining people's opinions have emerged as the most important part of social media analytics. The method of opinion mining in social media while dealing with various kinds of opinion documents as well as the issues connected with opinion mining from social media have been described. Specifically, this study focuses on social media. Twitter is a massive online social activity that allows millions of individuals to exchange their thoughts and ideas with one another. One key challenge is how to use sentiment analysis on the data collected from social media platforms to get product evaluations based on the product's characteristics. The K-means clustering approach was used in a sample Twitter dataset to cluster various attitudes in context with various attributes of goods. The results of this clustering were then evaluated and explained with the assistance of a machine-learning tool.

Zhao et al., (2014) [18] described that sentiment analysis tasks often need aspect identification and grouping. Numerous algorithms for identifying product features have been investigated so far, but there has been much less effort put into categorizing and grouping such features. It examines the issue of how to most effectively group similar product features. It is argued that two useful aspect relations must be used to characterize the connections between two aspects such as the relevant aspect relation and the irrelevant aspect relation. Aspect similarity calculation is performed to cluster the aspects into distinct groups using a hierarchical clustering technique using the relevant aspect set and the irrelevant aspect set for each product aspect. The experimental results on the camera domain demonstrate the efficacy of the two-aspect relations and the superior performance of the suggested method compared to the baseline without it.

1.2 Comparison of reviewed technique

The following tableanalyses the previous workson feature-based opinion classificationwith their findings described below:

Table1.Summary of Reviewed Techniques.

Authors [Ref.]	**Technique**	**Outcome**
Srivastava and Vijay et al., (2022) [11]	ML and lexicon-based algorithm	New automated methods are capable of solving all of the issues and producing the best possible outcomes.
Kumar et al., (2022) [12]	ML	Experiments were run with the aid of the Naive Bayes, FLR, and AdaBoost classifiers, and the results were compared to evaluate and assess the feature selection strategies.
Pradeepa et al., (2021) [13]	Helly property method	Experimental results demonstrate the efficacy of the suggested method for extracting aspect-sentiment, categorizing, and summarising evaluations of products posted on the web.
Alamoudi et al., (2021) [14]	ALBERT	The suggested approach for extracting aspects was successful to the tune of 83.04% of the time.
Ahamed et al., (2019) [15]	SVM and NB	It is also evaluated with state-of-the-art techniques, which offer an upper suggestion of 94.37% accuracy and reliability using Kappa indications displaying fewer error rates.

DOI: https://doi.org/10.17762/ijritcc.v10i2s.5914
***Article Received**: 12 October 2022 **Revised**: 18 November 2022 **Accepted**: 26 December 2022*

Zainuddin et al., (2018) [16]	PCA and RP	The results of the implemented systems demonstrated that the novel hybrid sentiment classification improved accuracy performance by 76.55, 71.62, and 74.24% over the previous baseline sentiment classification techniques.
Babu et al., (2017) [17]	K-means clustering approach	The results of this clustering were then evaluated and explained with the assistance of a machine-learning tool.
Zhao et al., (2014) [18]	Machine learning	The experimental findings on the camera domain show that the suggested technique outperforms the baseline without the two-aspect relations, and also show that the two-aspect relations are efficient.

III. Background study

Sentiment analysis must be performed using a set of criteria, such as whether the text is positive, negative, or neutral. In the wake of the meteoric rise of e-commerce over the last several decades, more and more companies are soliciting feedback from their satisfied customers. Due to the daily production of millions of evaluations, it is becoming more difficult for consumers to make informed purchases. Manufacturers have a tough time and spend a lot of time trying to analyse these massive ideas. The field of NLP has recently shifted its focus to deep learning. The suggested approach makes advantage of Skip-gram architecture to accurately glean semantic and contextual information from words in a corpus. The suggested model uses long short-term memory (LSTM) to comprehend intricate patterns in textual information. The LSTM's performance may be enhanced by using the adaptive particle Swarm Optimization method on the weight parameters. Extensive trials on four datasets show that the suggested APSO-LSTM model obtained more accuracy than the established approaches such as traditional LSTM, ANN, and SVM. Simulated results show that the suggested model outperforms the state-of-the-art in a variety of ways [19].

IV. Problem Formulation

People make decisions based on the emotions and perspectives of others around them. People's opinions and assessments of others are heavily influenced by their preconceived notions about other people. A user's opinion is an expression of his or her personal views on a subject. As a vital and rapidly developing field of study, Opinion Mining also presents a constantly renewing set of challenges. There is still a need for study and development into issues like spam filtering, determining the underlying phrase of an opinion, distinguishing between a valuable and useless opinion, and so on. Opinions are sometimes generated on the spot, and they may be vastly different from one person to the next, which can lead to several problems. While several methods exist for mining public opinion, each has its own set of problems. Authorization, non-expert opinion, spam opinion identification, opinion credibility, the relevance of the views, Natural Language Processing overhead, typographical mistakes, and so on are only a few of the many difficulties encountered in opinion mining.

- **Authorization:** The legitimacy of the point of view is of the utmost importance. The information source must be well-known within the community. In other words, the approval of the opinion holder may be decided by some different things, such as whether or not the opinion is offered by a certain domain or a compelling cause.
- **Non-expert Opinion:** The perspectives of non-experts cause havoc on websites and in public discussion forums. Opinion mining may be influenced by exceptionally focused websites such as programmer blog postings, in which the opinion of an expert is vital; nevertheless, these websites do not appropriately evaluate.
- **Natural Language Processing Overhead:** Natural language constraints such as uncertainty, co-reference, implicitness, interpretation, etc. hindered sentiment analysis methods.
- **Typographical Errors:** Words such as "knowledge for knowledge" and "a school for school" are comprehensible to humans, along with a variety of other grammatical errors. On the other hand, for an opinion miner, it might be challenging to extract and evaluate typo errors.

V. Research Methodology

The study model will highlight the numerous procedures associated with the various sorts of review analysis, such as dataset collecting (reviews), data pre-processing, and feature extraction. In the first step, collect the input review data and dataset taken from an online website or institution. Then pre-processing is applied to the given dataset. After that feature extraction will extract the semantic feature and

DOI: https://doi.org/10.17762/ijritcc.v10i2s.5914
***Article Received**: 12 October 2022 **Revised**: 18 November 2022 **Accepted**: 26 December 2022*

reliable optimized features set using Kernel Principal Component Analysis (KPCA). The CNN model is then initialized. A CNN model usually has three layers: a convolutional layer, a pooling layer, and a fully connected layer. The fully connected layer is the last layer. A kernel or filter is used in the convolution layers. This kernel or filter is located inside this layer and moves through the receptive fields of the image or object while testing to see whether a feature is present in the image or object, the pooling layer reduces the complexity of the CNN and improves its efficiency, and the full connected layer is where object classification happens based on the features extracted in the previous layers so that the classification of samples can be performed. The overall process is described in the next section in steps.

1.3 Techniques

In this methodology, there are used two methods named convolution neural network and kernel principal component analysis which are described below.

- **Convolution Neural Network (CNN)**

The architecture of a CNN is a neural network, which consists of numerous hidden layers, each of which consists of several two-dimensional planes that are populated by several neurons. In addition, it is considered that each neuron operates on its own. The feature extraction module is an integral part of the CNN architecture and may be thought of as a two-dimensional picture as the source of its input data. Figure 2 presents a schematic fundamental structure of CNN.

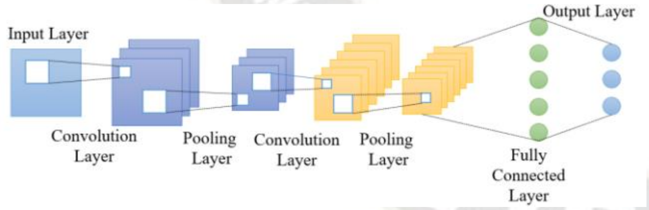

Figure2. The Basic Architecture of CNN [20].

Input Layer: The raw dataset that is being entered can be sent to the input layer in its original form. A single image is "inputted" into the input layer under the pixel value of the image.

Convolutional Layer: The up-sampling layer, whose purpose is to extract characteristics from the incoming data, comes by several different names. Different characteristics may be extracted from the input data by using convolutional layers, each of which has its convolutional kernel. As more up-sampling layers are used, more convolutional kernels may be used, resulting in a larger number of retrieved features.

Down-sampling Layer: It is also called as pooling layer. Its principal goal is to finalize the feature data extractions started in the convolutional layer. The baseline CNN design calls for a minimum of two convolutional layers and two down-sampling layers. Extracting characteristics from the input data is more likely to aid evident categorization as additional levels of the architecture are created.

Fully Connected Layer: Each of the feature maps serves as an individual input into the system. In a broad sense, the nodes of the neurons in the layer below are linked to the nodes of the layer above, but the nodes in each layer are not connected. This layer is responsible for integrating and normalizing the abstracted characteristics that were produced by the layers that came before it to provide a probability for the different situations.

Output Layer: The total number of neurons in this layer is determined by the requirements laid down. If the classification is necessary, the number of neurons would often be proportional to the number of categories that need to be categorized.

- **Kernel Principal Component Analysis**

Kernel principal component analysis is a method of non-linear dimensionality reduction. It is an extension of principal component analysis using kernel approaches, which is a technique for linear dimensionality reduction. A kernel function is used in kernel principal component analysis to project the dataset into a higher dimensional feature space where the dataset is linearly separable. There are many different kernel approaches, including linear, polynomial, and Gaussian methods. It provides a set of multi-dimensional evaluations that employ statistics on two levels and a small-scale group in which qualities are unrelated to both one and all [21].

Steps of kernel principal component analysis

1. First, we will select kernel functions k (x_i,x_j) and let Tbe any transformation to a higher dimension.

2. The data's covariance matrix would be determined thereafter. In contrast, we would now use kernel functions to arrive at this matrix's solution. This means we would calculate a matrix called the kernel matrix, which is obtained by applying the kernel function to each pairof data.

$$K = T(X)\,T(X)^T$$

3. Locate the centre of the kernel matrix (similar to taking the standard deviation of the modified data and dividing by the mean).

$$K_{new} = K – 2(I)K + (I)K(I)$$

Where I is a matrix with all elements equal to i/d.

4. The matrix's eigenvectors and eigenvalues will then be calculated.
5. Arrange the eigenvectors by their eigenvalues and sort the list in descending order.
6. For the simplified dataset, we'll choose some number of dimensions, which is called m. Next, we'll choose the first m eigenvectors and combine them into a single matrix.
7. The last step is to multiply the data using the matrix. New streamlined data set will be the result.

1.4 Dataset description

IMDB dataset consisting of 50,000 movie reviews, suitable for use in natural language processing or text analytics. The preceding benchmark datasets pale in comparison to this one, which is a dataset for binary sentiment classification and contains a significant amount more data than its predecessors. It gives a collection of 25,000 highly polar movie reviews that can be used for training, and it also provides 25,000 that can be used for assessment. Therefore, it predicts the amount of positive and negative reviews using either classification algorithms or deep learning techniques [22].

VI. Proposed methodology

The structure of the proposed methodology is shown in Figure 3. which is described further.

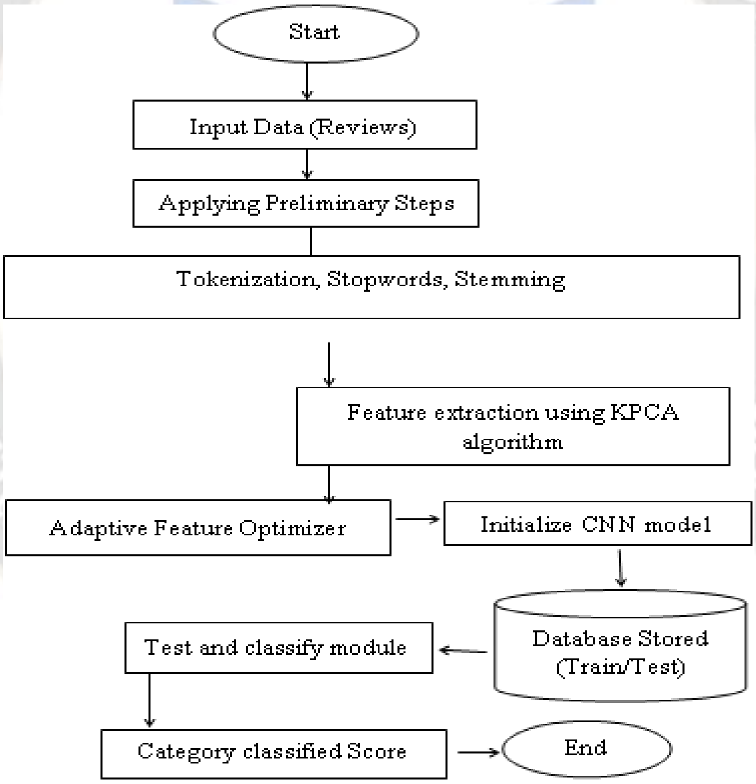
Start
Input Data (Reviews)
Applying Preliminary Steps
Tokenization, Stopwords, Stemming
Feature extraction using KPCA algorithm
Adaptive Feature Optimizer
Initialize CNN model
Database Stored (Train/Test)
Test and classify module
Category classified Score
End

Figure3. A proposed framework.

There are the following steps as described below in the above process.

Initial, it will collect the input review data with different categories such as positive, and negative, and the dataset is taken from an online website. After uploading the input review data, it identifies the feasibility of the dataset. In the second phase: pre-processing is applied to the dataset which are follow as:**Tokenization**: it is process reviews that are split into symbols, tokens, and words.**Stop word:** stop words such as I, and, for, the, and would. These words are removed from the NLP toolkit stop word list.**Stemming:** it overcomes the word length to its base forms.

In next phase, feature extraction is applied after pre-processing phase; it will extract a collection of features that are both semantically and statistically significant using kernel principal component analysis. The feature optimization procedure operates on the extracted feature vector data. This component handles the dataset, generates error-reducing feature vectors, and feeds them into classification models. In next phase, CNN is a type of neural network for processing data that is widely used for image/object recognition and classification is initialized. Then, testing and classification of the module from the stored database to achieve category classification score. Finally, several performance metrics are calculated and analysed.

VII. Results Discussion

Figure 4 depicts the 2x2 confusion matrix as can be seen; this matrix is described in terms of both its true value and its predicted value. There are total 12000 samples used in this matrix. In this figure, the 0 represents the negative, while the 1 represents the positive. Count values are used to provide a concise summary of the numbers of accurate and inaccurate predictions. 12323 is the representation of the true positive numbers, whereas the number 16 is the representation of the false positive numbers. The number 12 is used to symbolize a false negative, whereas the number 12327 is used to signify actual negatives. Figure 5 depicts the area under the curve (AUC), which is shown by the blue line in the figure that follows and is defined in terms of the ratio between the true positive rate and the false positive rate. The true positive rate is the percentage of observations that are expected to be positive when, in fact, they are positive. The false positive rate is the proportion of observations that are anticipated to be positive when, in fact, they are negative. The measureof area under the curve (AUC) is the 0.95. Figure 6 illustrates the model accuracy and the loss of training. There are five models, each of which is signified by a separate line in terms of the relationship between accuracy and epochs. Model accuracy was a representation of the model's performance, which was defined as a continuous improvement in a manner that was easy to understand and computed as a percentage. The loss is stated in terms of between loss and epochs, and its interpretation is defined as being that models are continuously going down in decreasing form. Loss is not given in terms of percentages. Figure 7 depicts the model accuracy and loss of validation. Five models are described in terms of between accuracy and epochs as well as loss and epochs. Model 1 attained the maximum accuracy which is 95.6% as compared to other models and also proposed technique achieved better Accuracy, Precision and F-measure as shown in Figure 8. Figure 9 is a demonstration of the Histogram graph, which represents the length of the document. This graph displays the distribution of a numeric variable's values as a series of bars for better visibility. Every individual bar, on average, contains a certain range of numerical values that is referred to as a bin. In this figure, the height of the bars indicates the number of rows contained in the document, while the width of the bars indicates the number of columns included in the document. Together, these two axes represent the total number of rows and columns in the document.

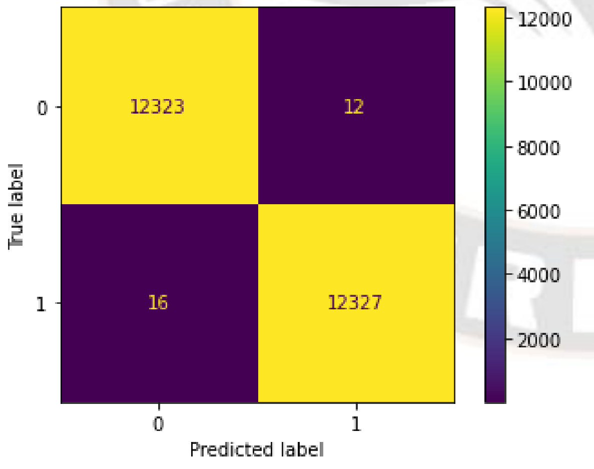

Figure4. Confusion matrix.

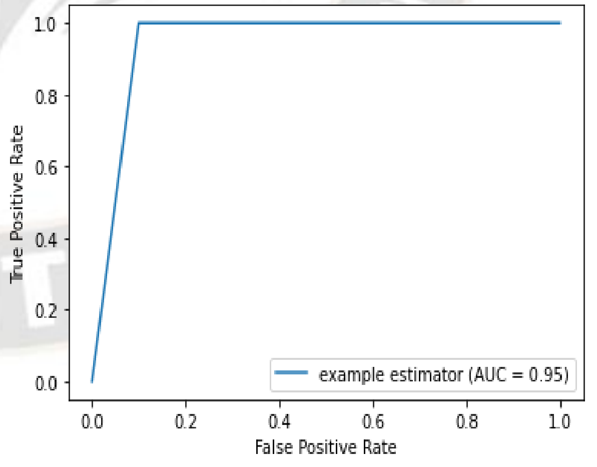

Figure5. AUC curve.

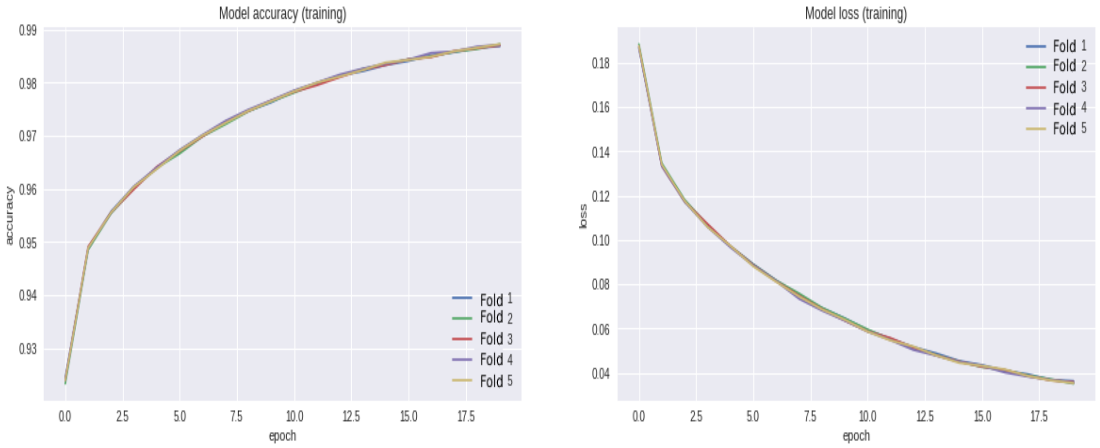

Figure6. Model Accuracy and Loss (Training).

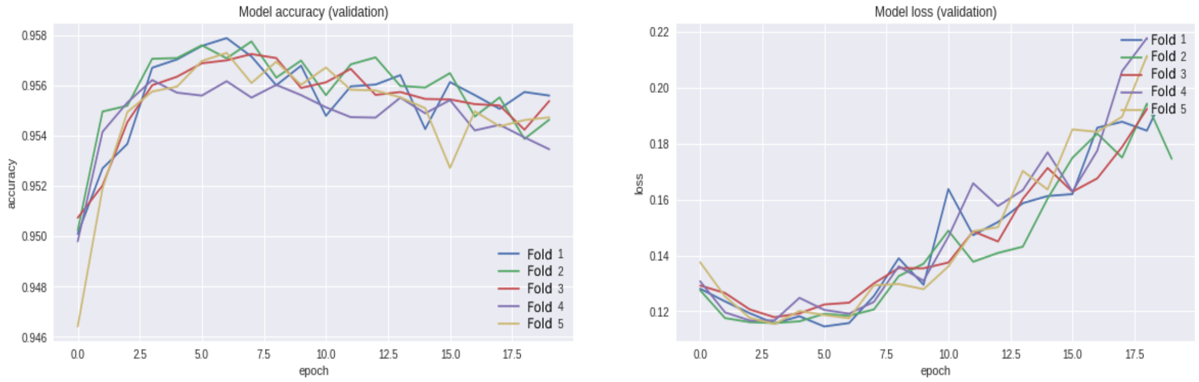

Figure7. Model Accuracy and Loss (Validation).

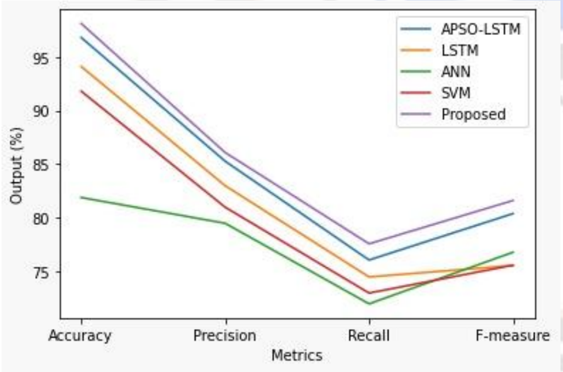

Figure8. Performance Evaluation

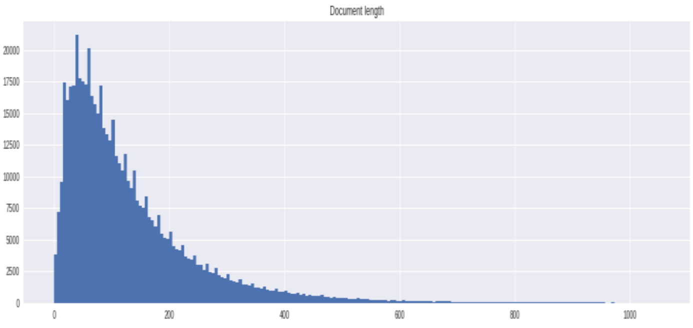

Figure9. Histogram graph.

VIII. Conclusion

In this study, a feature-based opinion categorization system that is both efficient and effective is developed. An additive CNN that uses the KPCA algorithm may tackle problems such as spam filtering, finding the underlying phase of an opinion, typographical mistakes, and discriminating between a worthwhile and a worthless opinion. Because of its strong capacity for generalization, KPCA method achieved good results. The dataset from the IMDB review was used for the simulation. It has been determined that the suggested approach is effective via its performance assessment. Following validation, the suggested approach achieved an accuracy level that was a maximum of 95.6%. The AUC curve has also been generated. The suggested technique has been shown to have better performance, as shown by empirical data.

References

[1]. Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies* 5, no. 1 (2012): 1-167.

[2]. Zhu, Jingbo, Huizhen Wang, Muhua Zhu, Benjamin K. Tsou, and Matthew Ma. "Aspect-based opinion polling from customer reviews." *IEEE Transactions on affective computing* 2, no. 1 (2011): 37-49.

DOI: https://doi.org/10.17762/ijritcc.v10i2s.5914
Article Received: 12 October 2022 Revised: 18 November 2022 Accepted: 26 December 2022

[3]. Medhat, Walaa, Ahmed Hassan, and HodaKorashy. "Sentiment analysis algorithms and applications: A survey." *Ain Shams engineering journal* 5, no. 4 (2014): 1093-1113.

[4]. Eirinaki, Magdalini, ShamitaPisal, and Japinder Singh. "Feature-based opinion mining and ranking." *Journal of Computer and System Sciences* 78, no. 4 (2012): 1175-1184.

[5]. Dubey Veena, G. D. "Sentiment Analysis Based on Opinion Classification Techniques: A Survey." *International Journal of Advanced Research in Computer Science and Software Engineering* (2016): 53-58.

[6]. Hemmatian, Fatemeh, and Mohammad Karim Sohrabi. "A survey on classification techniques for opinion mining and sentiment analysis." *Artificial intelligence review* 52, no. 3 (2019): 1495-1545.

[7]. Wójcik, Katarzyna, and JanuszTuchowski. "Comparison analysis of chosen approaches to sentiment analysis." *IT for practice* (2012): 187-192.

[8]. Wójcik, Katarzyna, and PawełWołoszyn. "Opinion Mining Framework and Its Applications." *CONTEMPORARY ISSUES IN ECONOMICS, BUSINESS, AND MANAGEMENT* (2016): 341.

[9]. Wójcik, K., and J. Tuchowski. "Sentiment Analysis of Opinions about Hotels Extracted from the Internet [in:] Knowledge–Economy–Society." *Global and Regional Challenges of the 21st Century Economy* (2013): 755-771.

[10]. Kontopoulos, Efstratios, Christos Berberidis, TheologosDergiades, and Nick Bassiliades. "Ontology-based sentiment analysis of Twitter posts." *Expert systems with applications* 40, no. 10 (2013): 4065-4074.

[11]. Srivastava, Durgesh, and Vijay Kumar Soni. "A Systematic Review On Sentiment Analysis Approaches." In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 01-06. IEEE, 2022.

[12]. Kumar, R. Satheesh, A. Francis Saviour Devaraj, M. Rajeswari, E. Golden Julie, Y. Harold Robinson, and Vimal Shanmuganathan. "Exploration of sentiment analysis and legitimate artistry for opinion mining." *Multimedia Tools and Applications* 81, no. 9 (2022): 11989-12004.

[13]. Pradeepa, S., N. Sasikaladevi, and K. R. Manjula. "Emotion aware feature based opening mining on large scale data by exploring hypergraph with Helly property." *Multimedia Tools and Applications* 80, no. 20 (2021): 30919-30938.

[14]. Alamoudi, Eman Saeed, and Norah Saleh Alghamdi. "Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings." *Journal of Decision Systems* 30, no. 2-3 (2021): 259-281.

[15]. Ahamed, Shoieb, AjitDanti, and S. P. Raghavendra. "Feature-Based Fuzzy Framework for Sentimental Analysis of Web Data." In *2019 International Conference on Data Science and Communication (IconDSC)*, pp. 1-5. IEEE, 2019.

[16]. Zainuddin, Nurulhuda, Ali Selamat, and Roliana Ibrahim. "Hybrid sentiment classification on Twitter aspect-based sentiment analysis." *Applied Intelligence* 48, no. 5 (2018): 1218-1232.

[17]. Babu, Anjan G., Surya S. Kumari, and K. Kamakshaiah. "An experimental analysis of clustering sentiments for opinion mining." In *Proceedings of the 2017 International Conference on Machine Learning and Soft Computing*, pp. 53-57. 2017.

[18]. Zhao, Yanyan, Bing Qin, and Ting Liu. "Clustering product aspects using two effective aspect relations for opinion mining." In *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, pp. 120-130. Springer, Cham, 2014.

[19]. Shobana, J., and M. Murali. "An efficient sentiment analysis methodology based on long short-term memory networks." *Complex & Intelligent Systems* 7, no. 5 (2021): 2485-2501.

[20]. O'Shea, Timothy, and Jakob Hoydis. "An introduction to deep learning for the physical layer." *IEEE Transactions on Cognitive Communications and Networking* 3, no. 4 (2017): 563-575.

[21]. Lima, Amaro, Heiga Zen, Yoshihiko Nankaku, ChiyomiMiyajima, Keiichi Tokuda, and Tadashi Kitamura. "On the use of kernel PCA for feature extraction in speech recognition." *IEICE TRANSACTIONS on Information and Systems* 87, no. 12 (2004): 2802-2811

[22]. IMDB Dataset of 50K Movie Reviews. (2022). Retrieved from https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews