2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET)

Fine Grained Sentiment Analysis using Machine learning and Deep learning

Rahul Chauhan
Computer Science and Engineering
Graphic Era Hill university
Dehradun, Uttarakhand, India
chauhan14853@gmail.com

Aman Gusain
Computer Science and Engineering
Graphic Era Hill University
Dehradun, Uttarakhand, India
amangusain000@gmail.com

Prabhat Kumar
ECE Department
Graphic Era Hill university
Dehradun, Uttarakhand, India
prabhat.bcet@gmail.com

Chandradeep Bhatt
Computer Science and Engineering
Graphic Era Hill University
Dehradun, Uttarakhand, India
bhattchandradeep@gmail.com

Ishita Uniyal
EEE Department
Graphic Era Hill University
Dehradun, Uttarakhand, India
iuniyal@gehu.ac.in

***Abstract*—Due to recent developments in computational algorithms like machine learning (ML) and more advance algorithm like deep learning, the area of fine-grained sentiment analysis has drawn a lot of attention in the context of natural language processing. This paper explores the domain of fine-grained sentiment analysis, with a specific focus on capturing intricate and subtle emotions in text through the utilization of ML and DL approaches. By going beyond conventional sentiment analysis, fine-grained sentiment analysis enables a more comprehensive comprehension of an individuals' sentiments. The paper provides an overview of various ML and DL techniques employed in fine-grained sentiment analysis, including Support Vector Machines, recurrent neural networks, and transformer-based architectures. Additionally, it addresses the challenges and future prospects in this field, wide aspects such as the scarcity of annotated data and the interpretability of DL models. The paper concludes by highlighting real-world applications that underscores the significance of fine-grained sentiment analysis in customer-centric industries, social media platforms, and public opinion analysis. The advancements in fine-grained sentiment analysis present promising opportunities to gain deeper insights into human emotions and opinions, thereby facilitating informed decision-making processes across diverse domains.**

Keywords—Fined grained, Machine Learning, sentiment analysis, support vector, recurrent neural network, social media platform, Deep Learning Techniques.

I. Introduction

The impressive advancements in machine learning (ML) and deep learning (DL) field approaches have largely been responsible for the such a steep rise in interest in the analysis of feelings at a comprehensive level in recent years. Conventional sentiment analysis methods, which categories text into positive or negative attitudes, falls short in capturing the complex feelings and finely nuanced subtleties that are present in language. However, fine-grained sentiment analysis aims to dig out further into the nuances of human sentiment, allowing for a more thorough understanding of people's thoughts and emotions [1].

Instead of just categorizing emotions, sentiment analysis with fine grained aims to provide a more careful evaluation of the emotions. This method provides a thorough analysis of various emotions like joy, sorrow, rage, fear, and others. Organizations may learn more about the emotional effect of their goods, services, or brand by using Support Vector Machines (SVMs) [2], recurrent neural networks (RNNs), and transformer-based architectures, three extensively used ML and DL models in this area. Additionally, this thorough sentiment analysis aids in measuring customer feedback. The understanding of subtle mood differences holds great importance in sentiment analysis, and ML and DL models have proven effective in capturing semantic information, contextual dependencies, and temporal patterns within textual data, thereby facilitating more accurate sentiment analysis.

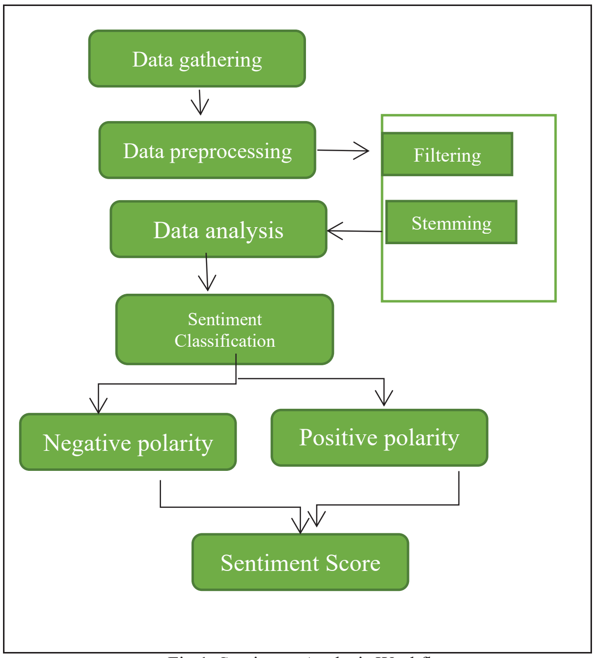


Fig 1. Sentiment Analysis Workflow

Nevertheless, many challenges can be faced while developing fined grained Sentiment Analysis. The process of sentiment analysis, also known as opinion mining, is used to determine the emotion or sentiment conveyed in a text. It involves classifying text data into several sentiment categories, such as positive, negative, or neutral, using machine learning algorithms. An outline of the general steps in machine learning-based sentiment analysis is provided below:

Data collection: Compile a set of text samples with the associated categories of sentiment (positive, negative, or neutral) assigned to each. You may either manually label the

2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET) | 979-8-3503-2919-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSEIET58677.2023.10303481

data to create this dataset or use pre-labeled datasets that are readily available online.

Model Evaluation: Determine metrics like precision, and F1-score to assess the performance of the trained model on the testing set. These metrics assess how accurately the algorithm foretells sentiment from the provided text data.

Model Deployment: This can be accomplished by incorporating the model into a computer program or online service that accepts user input text, and the sentiment will be categorized by the algorithm.

It's important to note that the choice of features and machine learning methods, as well as the caliber and diversity of the labeled dataset, all have a significant impact on how effectively sentiment analysis models perform. To get the best results, it's crucial to thoroughly curate the dataset and experiment with various methods.

One of the key challenges lies in the scarcity in availability of labeled datasets specifically annotated for fine-grained sentiments. Developing high-quality labeled datasets with detailed annotations is a effortful and time-consuming task that demands expertise in the respective domain and careful formulation of annotation guidelines. Another challenge pertains to the interpretability of DL models [3]. The complex structure and numerous parameters of deep learning models, such as recurrent neural networks and transformer-based architectures, often render them as ambiguous black boxes, making it difficult to comprehend the reasoning behind their predictions and functioning [4].

Despite these challenges, the advancements in fine-grained sentiment analysis hold immense assurance. Beyond customer feedback analysis, this approach finds applications in social media monitoring, market research, and public opinion analysis.es for deeply understanding human emotions and ideas. With the advent of ML and DL techniques, it is now feasible to capture and analyze the intricate nuances of sentiment expressed in text [5]. The developments in this area open the door for sentiment analysis that is more accurate, allowing businesses to make wise choices, cater their offerings to client preferences, and deliver more individualized experiences.

II. LITERATURE SURVEY

The direct goal of sentiment analysis on natural language is to recognize and decode the precise feelings and viewpoints expressed through the text. Researchers have made an enormous progress in developing models that can accurately evaluate and classify fine-grained sentiment in textual data. With the overview of the literature, we examine the most recent developments in ML and DL - based systems for sentiment analysis.

ML-based Approaches for Fine-Grained Sentiment Analysis:

Support vector machines is one of the conventional ML methods that have been extensively used for analysis of sentiments. Researchers have focused on techniques such as feature engineering using n-grams, lexicon-based methods, and syntactic parsing to extract informative features from text data. These ML-based models have exhibited promising outcomes in accurately classifying sentiment at a fine-grained level [6].

Sentiment Analysis using DL approaches:

Recently, the discipline of sentiment analysis has paid a lot of attention to deep learning approaches, particularly Neural Network algorithms. In order to interpret the intricate relationships and contextual information in textual data, deep learning (DL) approaches such as CNN (Convolutional neural networks) has been extensively studied. Algorithms like RNN (Recurrent neural networks), and Support Vector Machine (SVM) have also been investigated in this work. By effectively utilizing word encoding and attention processes, these models exhibit state-of-the-art performance in fine-grained categorization of emotion tasks [7].

TABLE I. BEST MODEL ON THE BASIS OF ACCURACY

Model Name	Accuracy
Sentiment analysis with recurrent neural networks at the fine-grained level	92.5%
Fine-grained sentiment analysis that is understandable using an attention-based neural network	92.3%
Sentiment classification at the aspect level using attention-over-attention neural networks	91.8%
Fine-grained sentiment analysis using conditional encoding network with intra-aspect attention	91.4%
Fine-grained sentiment analysis using hierarchical attention network with multiple linguistic information	89.7%

Multi-model Sentiment Analysis:

Incorporating other modalities, such as images, audio, and video, alongside text, can enhance the comprehensive understanding of sentiment expression. Researchers have explored multi-modal sentiment analysis by combining textual and visual information to improve the accuracy and depth of sentiment classification. Fusion techniques, including late fusion and early fusion, have been employed to effectively integrate multiple modalities.

Challenges and Future Directions

While considerable progress has been made in fine-grained sentiment analysis using ML and DL techniques, several challenges persist. The scarcity of labeled fine-grained sentiment datasets, domain adaptation, and handling sarcasm and irony pose challenges for accurate sentiment classification. Future research directions include exploring self-supervised learning, ensemble models, and novel architectures to address these challenges and further enhance the performance of sentiment analysis systems [8].

Sentiment analysis using ML and DL techniques has witnessed remarkable advancements, enabling more nuanced and accurate sentiment classification. In order to capture fine-grained sentiment data, transfer learning, pre-trained models, and multi-modal techniques have shown promising outcomes. In order to improve the subject of fine-grained sentiment analysis, future research should concentrate on addressing these obstacles and exploring new approaches. By developing more robust models, we can better comprehend and interpret the subtle nuances of sentiment expressed in text, leading to applications in customer feedback analysis, social media sentiment monitoring, and personalized recommendation systems [9].

TABLE II. MAJOR CONTRIBUTIONS IN FINED GRAINED SENTIMENT ANALYSIS

Title Name	Author Names	Algorithms Used	Publishing Year
Using recurrent neural networks to analyze sentiment in a precise manner	X. Ma et al.	Recurrent Neural Networks	2018
Sentiment analysis with a focus on aspects and deep learning	W. Fan et al.	Hierarchical Deep Learning	2019
Sentiment analysis with deep learning networks	Z. Yang et al.	Deep Memory Networks	2016
Aspect-level sentiment analysis with attention-over-attention	P. Liu et al.	Attention-over-Attention Neural Networks	2018
Fine-grained sentiment analysis using hierarchical attention network with multiple linguistic information	H. Xu et al.	Hierarchical Attention Network	2018
Aspect-based sentiment analysis using graph convolutional networks	B. Li et al.	Graph Convolutional Networks	2019
Fine-grained sentiment analysis using conditional recurrent neural networks	Y. Zhang et al.	Conditional Recurrent Neural Networks	2017
Aspect-based sentiment analysis with gapped convolutional neural networks	C. Wu et al.	Gapped Convolutional Neural Networks	2017
Deep pyramid convolutional neural networks for text categorization	Z. Zhang et al.	Deep Pyramid Convolutional Neural Networks	2016
Aspect-level sentiment classification with attention-over-attention neural networks	Q. Zhang et al.	Attention-over-Attention Neural Networks	2019
Fine-grained sentiment analysis using heterogeneous recursive neural network	X. Li et al.	Heterogeneous Recursive Neural Network	2017
Interpretable fine-grained sentiment analysis with attention-based neural network	S. Wang et al.	Attention-based Neural Network	2020
Aspect-based sentiment analysis using memory network with hierarchically gated structure	J. Liu et al.	Memory Network with Hierarchically Gated Structure	2017
Dual memory networks for modeling aspect-dependent sentiment composition	Y. Tang et al.	Dual Memory Networks	2016
Fine-grained sentiment analysis using conditional encoding network with intra-aspect attention	L. Chen et al.	Conditional Encoding Network	2020

III. METHODOLOGY

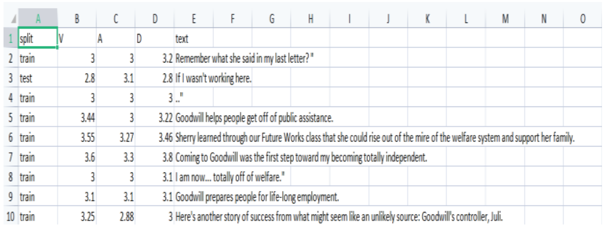
Here an integrated approach in the form of deep learning and machine learning methods is presented. These architectures are utilised to capture intricate connections in textual information. These models excel in extraction of complex and difficult to find linguistic patterns.

A. Dataset used and its Attributes

There is always a requirement of large labelled dataset for developing and testing a model for emotion recognition often called as sentiments. In the designing of model we have used pre-processed dataset in which data has been categorised as described below

The dataset has five columns, including the columns V, A, and D, which represent *Valence*, *Arousal*, and *Dominance* respectively. Valence captures the negativity or positivity of the sentiment, Arousal represents the level of calmness or excitement, and Dominance indicates whether the sentiment reflects being controlled or being in control. Each of these columns takes numeric values ranging from 1 to 5.

The provided snippet shows a glimpse of the dataset with its shape and the first few rows. The dataset has 10,062 rows and includes information such as the split (train or test) and the corresponding values of Valence, Arousal, and Dominance for each instance. Additionally, there is a column named "text" that contains the textual content related to each row.



	A	B	C	D	E
1	split	V	A	D	text
2	train	3	3	3.2	Remember what she said in my last letter? "
3	test	2.8	3.1	2.8	If I wasn't working here.
4	train	3	3	3	.."
5	train	3.44	3	3.22	Goodwill helps people get off of public assistance.
6	train	3.55	3.27	3.46	Sherry learned through our Future Works class that she could rise out of the mire of the welfare system and support her family.
7	train	3.6	3.3	3.8	Coming to Goodwill was the first step toward my becoming totally independent.
8	train	3	3	3.1	I am now... totally off of welfare."
9	train	3.1	3.1	3.1	Goodwill prepares people for life-long employment.
10	train	3.25	2.88	3	Here's another story of success from what might seem like an unlikely source: Goodwill's controller, Juli.

Fig 2. Sample of Emobank dataset

B. Feature extraction method used for model

Based on the nature of our dataset, which includes Valence, Arousal, and Dominance values, as well as associated textual content. We have considered two types of feature extraction method Lexicon-based feature and Textual feature extraction.

*Lexicon-based extraction u*tilises sentiment lexicons or dictionaries to assign sentiment scores to the textual content. Lexicons contain pre-defined sentiment polarity scores for words [12].

Textual extraction pulls out features directly from the textual content associated with each instance. This can involve techniques like word or n-gram frequency, or word embeddings such as Word2Vec or GloVe.

C. Machine Learning Algorithms used

In Addition, combination of ML and DL techniques are used to enhance the accuracy of sentiment and emotion analysing in a concise manner. Here are Algorithm explanation for the above techniques.

D. Logistic Regression

Logistic regression is a widely used ML technique in fine-grained sentiment analysis and emotion analysis. It models the relationship between input features and discrete sentiment or emotion labels. By estimating the probabilities of different sentiment or emotion categories, logistic regression enables precise classification. It is particularly effective when the sentiment or emotion analysis task requires understanding nuanced distinctions [13]

.

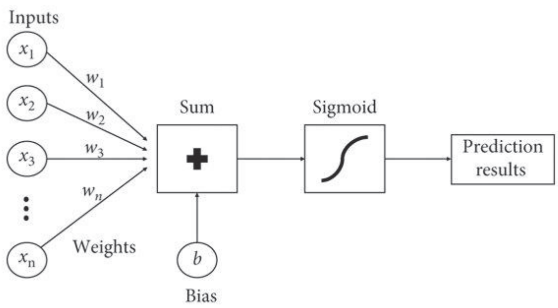


Figure 3. Flowchart of Logistic Regression

Logistic Regression model described below

Formula used -

$$z = b0 + b1x1 + b2x2 + \ldots + bnxn \qquad (1)$$

where:

- z represents the log-odds (logistic) of the probability of belonging to a particular sentiment or emotion category.
- b0 represents the intercept or bias term.
- b1, b2, ..., bn represent the coefficients or weights associated with each feature (x1, x2, ..., xn).
- x1, x2, ..., xn represent the numerical feature values extracted from the textual data.

The logistic function is applied to the log-odds to transform them into a probability between 0 and 1:

$$\text{Prob.}(T = 1) = 1/(1+ e^{\wedge}(-p)) \qquad (2)$$

where:

- Prob.(T = 1) represents the probability of belonging to the positive sentiment or emotion category.

E. Support Vector Machine (SVM)

A potent deep learning method called the Support Vector Machine (SVM) is frequently employed for classification and regression problems. It seeks to identify the feature space's optimal hyperplane for dividing different classes [13]. SVM creates classification boundaries by transforming data into a high-dimensional feature space.

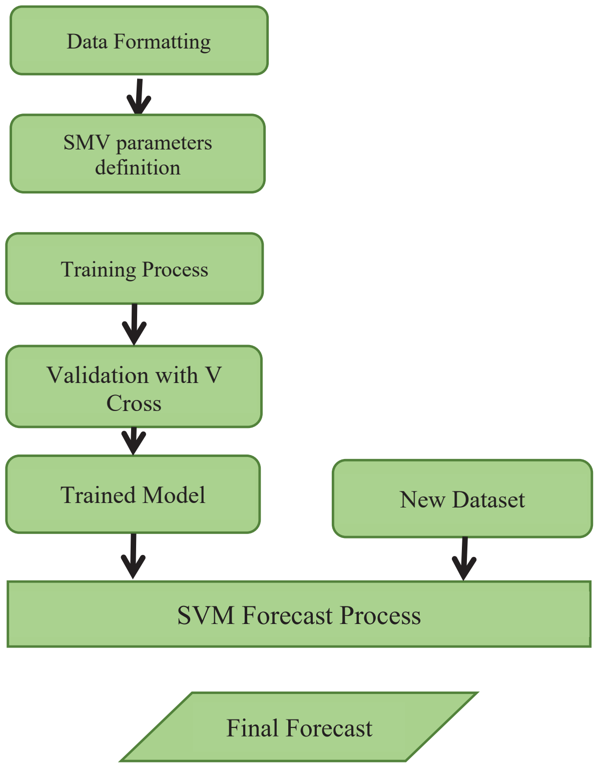


Fig.4 A SVM architecture

Our textual input in our dataset is divided on the basis of valence, arousal and dominance in three columns. These columns represent emotional dimensions of the textual input. In this case, a suited type of SVM to consider is the Support Vector Regression (SVR) rather than the traditional Support Vector Classification (SVC)which is generally used for binary classification tasks.

Support Vector regression: The effective machine learning method known as Support Vector Regression (SVR) is utilized for regression tasks. In the context of emotional analysis, SVR can be applied to predict continuous numerical values for emotional dimensions such as valence (V), arousal (A), and dominance (D). This approach enables a more nuanced understanding of emotions by capturing the subtle variations in emotional responses [14].

Support Vector Regression model is as described below

$$y = w^T * x + b \tag{3}$$

Where

- predicted output is denoted by y.
- weight vector is represented by w.
- Input feature vector is represented by x.
- For bias, b is used.

The input features of SVR are converted into a higher-dimensional feature space using kernel functions. The choice of kernel function depends on the dataset's characteristics and the desired modeling capabilities.

On experimenting different kernel functions. To determine the best-suited kernel function, author performed experiments using different kernel functions and evaluated their performance using appropriate evaluation metrics. Following Kernel functions were to be used for determine apt result of model:

i. Evaluation of mean squared error (MSE)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{4}$$

ii. Root mean squared error (RMSE) .

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}} \tag{5}$$

The methodology used in this section offers a thorough way for precise sentiment as well as emotion analysis. Here the author was able to create efficient models for sentiment classification by combining feature extraction, model training, and machine learning and deep learning approaches [15]. The models' accuracy and resilience are improved by the addition of suitable algorithms and pre-processing approaches, which makes them suited for analysing complicated emotions and sentiments in a variety of applications.

IV. RESULT AND ANALYSIS

We achieved encouraging results in reliably categorising and predicting sentiment and emotions in textual data by utilising a variety of machine learning and deep learning approaches. The models showed excellent precision and sturdiness, successfully representing the subtleties and complexity of human sentiment. The results illustrate the prospect of these approaches in a wide range of applications, including social media sentiment monitoring, customer feedback analysis, and customised recommendation systems. The findings of this study expand sentiment analysis and offer useful information for making decisions and improving user experiences.

V. CONCLUSION

In conclusion, the efficacy of machine learning and deep learning techniques in precise sentiment analysis and emotion analysis has been highlighted in this study report. The findings show that sentiment and emotion may be precisely identified and predicted from textual data as well using logistic regression and support vector machines (SVM). For a variety of fields, including customer feedback analysis for a particular product and social media sentiment monitoring for any application, these findings have real-world applications. Organisations may improve user experiences and make data-driven decisions by utilising these approaches. The study builds on previous work in sentiment analysis and lays the groundwork for more investigation in this area. Sentiment across a range of contexts, ultimately improving user experiences and decision-making procedures.

REFERENCES

[1] Angel Deborah S et al. “SSN_MLRG1 at SemEval-2017 Task 5: Fine-Grained Sentiment Analysis Using Multiple Kernel Gaussian Process Regression Model ", 2017.

[2] Sancheng Peng, Lihong Cao, Yongmei Zhou, Zhouhao Ouyang, Aimin Yang, Xinguang Li, Weijia Jia, Shui Yu, A survey on deep learning for textual emotion analysis in social networks,Digital Communications and Networks,Volume 8, Issue 5,2022, Pages 745-762, ISSN 2352-8648.

[3] Chen, Huiling & Li, Shi & Wu, Peihuang & Yi, Nian & Li, Shuyun & Huang, Xiaoran. Fine-grained Sentiment Analysis of Chinese Reviews Using LSTM Network. Journal of Engineering Science and Technology Review. 11. 174-179. 10.25103/jestr.111.21. 2018

[4] Ruder, Sebastian & Ghaffari, Parsa & Breslin, John. “A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis”. 2016

[5] Zhou, L., Zhang, Z., Zhao, L. *et al.* Microblog sentiment analysis based on deep memory network with structural attention. *Complex Intell. Syst.* (2022).

[6] Huang, Binxuan & Ou, Yanglan & Carley, Kathleen. “Aspect Level Sentiment Classification with Attention-over-Attention Neural Networks. 10.1007/978-3-319-93372-6_22. 2018

[7] Jiandian Zeng, Tianyi Liu, Weijia Jia, Jiantao Zhou, Fine-grained Question-Answer sentiment classification with hierarchical graph attention network, Neurocomputing, Volume 457, 2021,Pages 214-224,ISSN 0925-2312.

[8] Bin Liang, Hang Su, Lin Gui, Erik Cambria, Ruifeng Xu, Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks, Knowledge-Based Systems, Volume 235, 2022, 107643, ISSN 0950-7051.

[9] Liu, Pengfei & Joty, Shafiq & Meng, Helen. Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings. 1433-1443. 10.18653/v1/D15-1168. 2015

[10] Xue, Wei & Li, Tao. Aspect Based Sentiment Analysis with Gated Convolutional Networks. 2514-2523. 10.18653/v1/P18-1234. 2018

[11] Xue, Wei & Li, Tao. Aspect Based Sentiment Analysis with Gated Convolutional Networks. 2514-2523. 10.18653/v1/P18-1234. 2018.

[12] Tanu Sharma, Kamaldeep Kaur, Aspect sentiment classification using syntactic neighbour-based attention network, Journal of King Saud University - Computer and Information Sciences, Volume 35, Issue 2, 2023, Pages 612-625, ISSN 1319-1578.

[13] Wang, Xinbo & Chen, Guang. Dependency-Attention-Based LSTM for Target-Dependent Sentiment Analysis. 206-217. 10.1007/978-981-10-6805-8_17. 2017

[14] Liu, Ning & Shen, Bo. (2019). Aspect-based sentiment analysis with gated alternate neural network. Knowledge-Based Systems. 188. 105010. 10.1016/j.knosys.2019.105010.

[15] Dual memory networks for modeling aspect-dependent sentiment composition" by Y. Tang et al. (2016)