EAI Endorsed Transactions
on Internet of Things
Research Article EAI.EU

Evaluation of Machine Learning Techniques for Enhancing Scholarship Schemes Using Artificial Emotional Intelligence

P S Raju[1*], Sanjay Kumar Patra[2], Binaya Kumar Patra[3]

[1,2,3]Computer Science Engineering & Applications, Indira Gandhi Institute of Technology (IGIT), Saranga, Odisha, India

Abstract

This paper investigates the sentiment analysis of the" scholarship system" [4], in Odisha, primarily, to identify why some students do not apply for government-sponsored scholarships. Our research focuses on social media platforms, surveys, and machine learning-based analyses to understand the decision-making process and increase awareness about the various scholarship schemes. The goal of our experiment is to determine the efficacy of sentiment analysis in evaluating the effectiveness of different scholarship schemes. A wide variety of techniques based on dictionaries; corpora lexicons are used in different scholarship schemes for sentiment analysis. Our research paper is based on an evaluation process that could have a positive effect on the government by improving scholarship programs and giving financial aid to students from poor families, which would raise the level of education in Odisha. Our research paper concludes with a summary of successful and unsuccessful schemes, as well as their Word frequency counts and Sentiment Polarity scores.

Keywords: Emotion AI and Opinion Mining, SentiwordNet, WordNet, and Sentiment Polarity Classification

Received on 15 December 2023, accepted on 03 March 2024, published on 11 March 2024

doi: 10.4108/eetiot.5368

1. Introduction

" Sentiment analysis is the process of extracting text for sentiment and subjective information such as whether it expresses good, negative, or neutral feelings using data mining, machine learning (ML), artificial intelligence, and computational linguistics" [10]. Evaluation research improves knowledge and decision-making, leading to practical applications. This paper looks at the importance of analyzing user-generated feelings through a case study of twelve Odisha Government Education Schemes. Focus on" sentiment analysis" [5], this research outlines the difficulties of understanding the positive and negative perspectives of a given word, phrase, sentence, text, etc. In order to help people, industry, government, and analysts in their decision-making, two forms of machine learning techniques will be used: supervised and unsupervised. With" unsupervised machine learning" [7], this paper seeks to gain a deeper understanding the mood analysis.

Ghosh et al.'s 2023 study on machine learning for [8] water quality analysis, 'Water Quality Assessment Through Predictive Machine Learning', explores predictive analytics for water parameters. In 2023, Rahat and Ghosh's 'Unraveling the Heterogeneity [9] of Lower-Grade Gliomas' discusses the use of deep learning in brain MR image analysis for medical insights. The 2023 work [10] by Ghosh, Rahat, and their team, 'Potato Leaf Disease Recognition and Prediction using Convolutional Neural Networks', demonstrates the use of neural networks in detecting potato leaf diseases. Mandava, Vinta, Ghosh, and Rahat's 2023 research, 'An All-Inclusive Machine Learning and Deep [11] Learning Method for Forecasting Cardiovascular Disease in Bangladeshi Population', integrates [12] AI for health forecasting. The study 'Identification and Categorization of Yellow Rust Infection

*Corresponding author. Email: pmunna2222@gmail.com

in Wheat through Deep Learning Techniques' by Mandava et al. in 2023, applies deep learning to wheat disease detection. Khasim, Rahat, Ghosh, and others' 2023 article, 'Using Deep [`13] Learning and Machine Learning: Real-Time Discernment and Diagnostics of Rice-Leaf Diseases in Bangladesh', explores AI in rice-leaf disease diagnosis. In 2023, Khasim, Ghosh, Rahat [14] and colleagues' 'Deciphering Microorganisms through Intelligent Image Recognition' discusses machine learning for microorganism identification. Mohanty, Ghosh, Rahat, and Reddy's 2023 study, 'Advanced [15] Deep Learning Models for Corn Leaf Disease Classification', focuses on deep learning for classifying corn leaf diseases. Alenezi and team's 2021 research [16]'Block-Greedy and CNN Based Underwater Image Dehazing for Novel Depth Estimation and Optimal Ambient Light' investigates CNN methods for underwater image enhancement.

1.1 ” Levels of Sentiment Analysis”

•” **Document-level Sentiment Analysis**:” [8][1] At this stage, the entire text is broken down into its component parts and the prevailing tone is identified. Feedback on the scholarship program is necessary.

• **Sentence Level Sentiment Analysis**: At this level, the sentiment analysis task searches each sentence and verifies the sentence as whether or not a positive, negative, or neutral opinion. Perspicacity Classification distinguishes between objective and subjective sentences, referred to as Sentence Level sentiment analysis.

• **Aspect Level Sentiment Analysis:” Aspect** Level Sentiment Analysis” [9][3], was earlier referred to as Feature Level Sentiment Analysis. Document and Sentence Level Sentiment Analysis is a finer-grained analysis that looks directly at the opinion rather than documents, paragraphs, sentences, clauses, or phrases. This paper uses Sentence Level and Document Level Sentiment Analysis to gain a deeper understanding of sentiment analysis.

2. Data Source

Government analyses people’s opinions individually to know about their scheme’s popularity, which affects the scheme’s quality.

• **Blogs:** Government can use blogs to communicate with citizens and Make decisions based on their emotions.

• **Data Set:** The government gathers information from original sources secondary sources as well as Govt. schemes data (https://data.gov.in/policies).Our research paper collected this dataset of 12 scholarship schemes using Google Forms (i.e. Google Forms).

3. Experimental Setup

The main objective of this experiment is to evaluate the effectiveness of different scholarship programs of the Odisha Government using sentiment analysis. A wide variety of techniques based on dictionaries, corpora other reference materials. The use of Google Forms, i.e. Google Forms, was made, in order to collect this data. Sentiment analysis was used to evaluate Odisha government scholarship programs, such as PRERNA SCHOLARSHIPS. The goal of our experiment is to determine the efficacy of sentiment analysis in evaluating the effectiveness of different scholarship programs.

3.1 Sentiment Analysis Process:

The” sentiment analysis” [10], can be achieved by two approaches i.e., supervised learning and unsupervised learning in sentiment analysis.

a. Data Collection; After collecting people’s opinions on various schemes using Google Forms, it can be organized, semi-structured, or unstructured.

b. Pre-processing: Data pre-processing is a technique, which removes incomplete noisy, and inconsistent data using the following steps: •Removing URLs, Special characters, Numbers, Punctuations, etc. •Removing Stopwords •Stem-ming •Tokenization.

c. Feature Selection: The motive of feature selection is to minimize the dimensionality of the feature space and computational cost.

d. Sentiment Word Identification: Word recognition and sentiment analysis using vectorizer function and this paper uses K-Means Clustering unsupervised learning algorithm to combine positive, negative, and neutral word sentiment categories.

e. Sentiment Polarity Identification: Sentiment analysis involves classifying opinion features into three categories: Positive, Negative, and Neutral. Polarity identification uses different lexicons. SentWordNet, etc.

f. Sentiment Classification: The sentiment classification of government schemes opinions dataset uses the Word frequency counts method and Sentiment Polarity scores.

g. Analysis of opinions: Government schemes are accepted if they are positive.

3.2 Sentiment Analysis process Approach

3.3 Sentiment Classification Approaches

4. Proposed Methodology

The Proposed methodology is the “Mixed-Sentiment Approach”, which must be utilized for mining the Google Form dataset and the processes that follow:

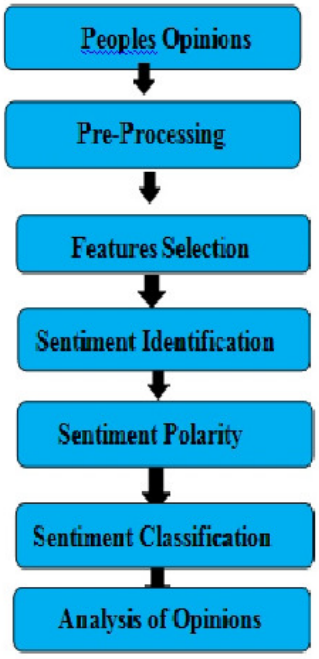

Figure 1. Sentiment Analysis Process Approach

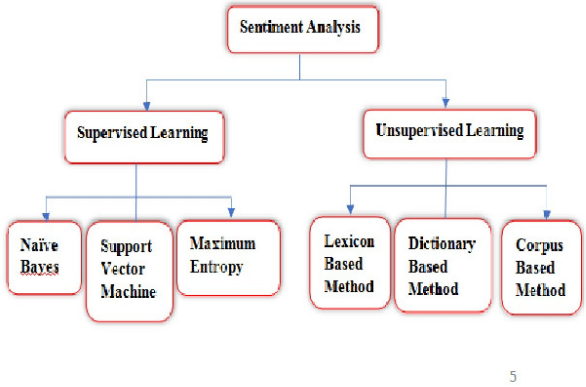

Figure 2. Sentiment Classification Approaches, it contains two methods i.e. Supervised Learning and Unsupervised Learning Approach

Sentiment Analysis process (algorithmic approach):

STEP-1: Data Collection of 12 schemes,
STEP-2- Pre-processing (Removing URLs, Special characters, Numbers, Punctuations),
STEP-3- Feature Selection (Removing Unwanted Data Fields),
STEP-4- Sentiment Word Identification,
STEP-5- Proposed Model:(Mixed Sentiment Analysis Process),
STEP-5.1: Corpus-based model,
STEP-5.2: Dictionary based model,
STEP-5.3: Lexicon based model,
STEP-6- Sentiment Classification,
STEP-7- Review-Results conclusion

4.1 Unsupervised Learning Approach

Using training data, unsupervised learning classifies sentiment. Domain dependency solves the problem and lowers training data. PMI is used to find seed words in a sentence. These terms allow us to calculate the average document sentiment.

1) Corpus Based Methods: In this paper, we have applied, the Corpus-Based Method that converts sentences to tokens, we converted the collected opinions to transformation.

a. Darmstadt service review corpus: An opinion sentiment corpus of 5513 hand-classified texts.

2) Dictionary Based Methods:

b. WordNet: In this paper, we have applied WordNet to discover the meaning, synonyms, and antonyms. In polarity of a word is determined by measuring its shortest distance between "good" and "bad". WordNet organizes English words into cognitive synonyms for research identification.

c. General Inquirer: General Inquirer (GI) is the oldest manual lexicon developed in 1966, used by scientists, political scientist's psychologists to identify characteristics of messages.

d. LIWC: LIWC is text analysis software that uses a proprietary dictionary of 4500 words in 76 categories to analyze emotional, cognitive, structural, and process components in text.

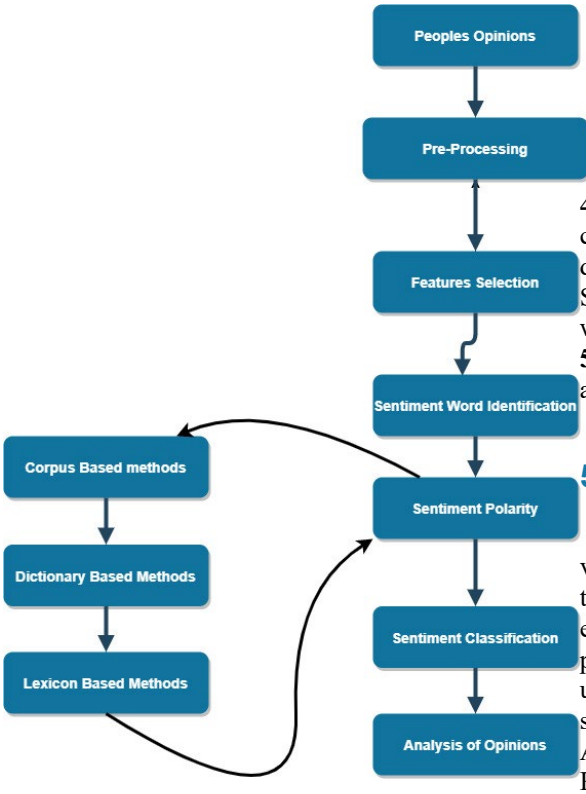

Figure 3. Mixed-Sentiment Approach

e. AFINN: AFINN is a list of English words manually labelled by Finn Arup Nielsen with an integer ranging from -5 to +5.

3) Lexicon-based Methods:

The lexicon-based method maps words to categorical or numeric scores based on their polarity in the text to determine the overall sentiment. There are several available lexical resources for this purpose.

f. SentiWordNet: In this paper, we have applied SentiWordNet to find out Document polarity scores using SentiWordNet for Positive and Negative Terms in documents with polarity scores. such as an example: 'good' has a Positive score:0.5, and 'good' has a Negative score:0.0." SentiWordNet" provides three statistical sentiment scores (positivity, negativity, neutrality) for each WordNet synset. A Word Sense Disambiguation (WSD) algorithm determines the most likely meaning.

g. WordNet-Affect:" WordNet-Affect" [6], is a linguistic resource that represents affective knowledge as an extension of WordNet. It uses a domain independent hierarchy to map the positive emotion label to the noun" illness".

h. MPQA: The resource provides a list of 8,222 terms labelled as subjective expressions, with polarity and intensity information.

i. SenticNet: Semantic computing is used to assign sentiment scores to 14,000 commonly understood concepts.

j. Hu and Liu's lexicon: Hu and Liu's lexicon is an opinion resource that lists positive and negative words in English.

4)Sentiment Classification: Sentiment Word classification of government schemes opinions dataset is done using sentiment Polarity scores from Mixed Sentiment Approach and Word frequency counts method, which uses a frequency distribution for experiment results.

5)Analysis of opinions: Government schemes are accepted if they are Positive Scores.

5. Experiment and Result

This study looks at three different tasks that various schemes performed: (1) opinion mining schemes that are expressed by people (2) Examine public opinion on each scheme and determine whether the sentence sense is positive or negative. (3) After aggregating the results, it is used to perform word processing to determine the sentiment. It had been used Unsupervised Learning Approaches which contain lexicons in this methodology. Related to sentiment analysis and opinion mining in this methodology lexicon-based methods have been used.

	Schemes	Tweets_Used	Year
0	PRERANA_SCHOLARSHIP	250	2020
1	E_MEDHABRUTI_SCHOLARSHIP	137	2020
2	ELEMENTARY_LEVEL_SCHEME_UP_MERIT	243	2020
3	ELEMENTARY_LEVEL_SCHEME_LP_MERIT	220	2020
4	SECONDARY_LEVEL_NTSE	251	2020
5	SECONDARY_LEVEL_NATIONAL_MEANS_CUM_MERIT	250	2020
6	SECONDARY_LEVEL_AMERICAN_INDIA_FOUNDATION_SCHO...	250	2020
7	PROFESSOR_GHANASHYAM_DASH_SCHOLARSHIP	250	2020
8	KALIA_SCHOLARSHIP	250	2020
9	PATHANI_SAMANTA_MATHEMATICS_SCHOLARSHIP	242	2020
10	CENTRAL_SECTOR_SCHOLARSHIP	250	2020
11	PM_SCHOLARSHIP_SCHEME	250	2020

Figure 4. Google Form Dataset Used for Experimental Work. In, Figure 4 shows the Word Cloud of Government Schemes

For calculating the Score of each opinion, we need the opinion lexicons and scoring function. The working of the scoring function is as follows:

Sentiment Score = Positive Words – Negative Words when Positive Polarity exceeds Negative Polarity, the score is positive. If Negative Polarity is greater than Positive Polarity, the score will be negative. If there are equal numbers of positive and negative words or no opinion words, the score is neutral. Figure 5 shows the Sentiment Analysis's distribution of Positive, Negative, and Neutral opinions of government schemes. This indicates how polarity varies with different Government Schemes on Google Form Dataset.

6. Conclusion

In conclusion, the mixed-sentiment approach has proven effective in analyzing government schemes. For 12 schemes, the proposed Mixed-sentiment Approach Result Graph shows higher word frequency count scores and evaluation results (Fig. 6). In the algorithmic approach, datasets are inputs, experiment outputs are either successful or unsuccessful schemes, and scheme opinions are collected, Unambiguity in nature, sentiment analysis (algorithmic approach) instructions are finite, effective, and language independent. It has been found that successful schemes are Prerana Scholarship, e-medhabruti scholarship, secondary level NTSE, secondary level national means cum merit, secondary level American India foundation scholarship, professor Ghanshyam dash scholarship, Kalia scholarship, Pathani Samanta mathematics scholarship, central sector scholarship, PM scholarship. Whereas the elementary level UP merit scheme and the elementary level LP merit scheme are unsuccessful due to a lack of awareness, short duration, and non-inclusivity of all the caste categories.

	Schemes	Positive_Sentiments	Neutral_Sentiments	Negative_Sentiments	Tweets_Used
0	PRERANA_SCHOLARSHIP	145	81	24	250
1	E_MEDHABRUTI_SCHOLARSHIP	53	60	24	137
2	ELEMENTARY_LEVEL_SCHEME_UP_MERIT	76	85	82	243
3	ELEMENTARY_LEVEL_SCHEME_LP_MERIT	64	69	87	220
4	SECONDARY_LEVEL_NTSE	85	117	49	251
5	SECONDARY_LEVEL_NATIONAL_MEANS_CUM_MERIT	110	103	37	250
6	SECONDARY_LEVEL_AMERICAN_INDIA_FOUNDATION_SCHO...	131	86	33	250
7	PROFESSOR_GHANASHYAM_DASH_SCHOLARSHIP	122	92	36	250
8	KALIA_SCHOLARSHIP	99	103	48	250
9	PATHANI_SAMANTA_MATHEMATICS_SCHOLARSHIP	118	106	26	242
10	CENTRAL_SECTOR_SCHOLARSHIP	140	92	18	250
11	PM_SCHOLARSHIP_SCHEME	104	115	31	250

Figure 5. Google Form Dataset Used for Experimental Work

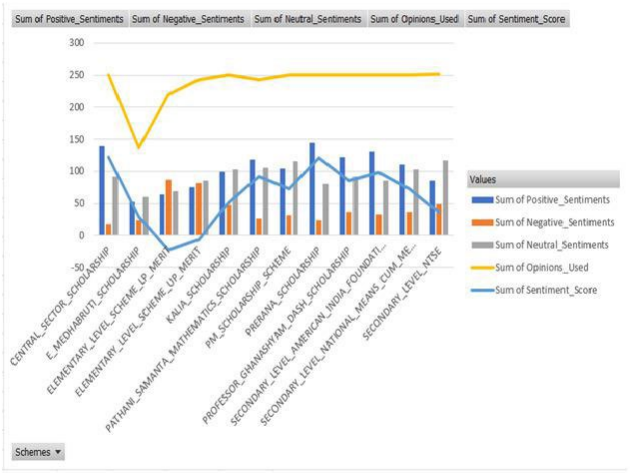

Figure 6. Google Form Dataset Used for Experimental Work

7. Future Work

Our future goal is to use sentiment analysis to evaluate scholarship programs. Machine learning and sentiment analysis are used to determine people's feelings. This research paper focuses on measuring public opinion of government programs, particularly scholarship programs. This system should make government decisions and scholarship program satisfaction easier.

8. References

[1] S. Behdenna, F. Barigou, and G. Belalem, "Document Level Sentiment Analysis: A survey," EAI Endorsed Trans. Context. Syst. Appl., 2018, vol. 4, no.13, p 154339, doi: 10.4108/eai.14-3-2018.154339.

[2] E. Cambria, "An introduction to concept-level sentiment analysis," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2013, vol. 8266 LNAI, no. PART 2, pp. 478–483, doi: 10.1007/978-3-642-45111- 941.

[3] H. Liu, I. Chatterjee, M. Zhou, X. S. Lu, and A. Abusorrah, "Aspect-Based Sentiment Analysis: A Survey of Deep Learning Methods", IEEE Trans. Comput. Soc. Syst., 2020, vol. 7, no. 6, pp. 1358–1375, doi: 10.1109/TCSS.2020.3033302.

[4] X. Ma, Q. Zhang, Y. Cui, J. Qu, and F. Yue, "Evaluation analysis of university student scholarship," Proc. Int. Symp. Test Meas., 2009, vol. 2, pp. 152–155, doi: 10.1109/ICTM.2009.5413089.

[5] S. Ranathunga and I. U. Liyanage, "Sentiment Analysis of Sinhala News Comments," ACM Trans. Asian Low-Resource Lang. Inf. Process., 2021, vol. 20, no. 4, doi: 10.1145/3445035.

[6] S. Poria, A. Gelbukh, E. Cambria, P. Yang, A. Hussain, and T. Durrani, "Merging SenticNet and WordNet-Affect emotion lists for sentiment analysis," Int. Conf. Signal Process. Proceedings, ICSP, 2012, vol. 2, pp. 1251–1255, doi: 10.1109/ICoSP.2012.6491803.

[7] R. Yang, "Unsupervised machine learning for physical concepts", 2022, pp. 31–34, [Online]. Available: http://arxiv.org/abs/2205.05279.

[8] Ghosh, H., Tusher, M.A., Rahat, I.S., Khasim, S., Mohanty, S.N. (2023). Water Quality Assessment Through Predictive Machine Learning. In: Intelligent Computing and Networking. IC-ICN 2023. Lecture Notes in Networks and Systems, vol 699. Springer, Singapore. https://doi.org/10.1007/978-981-99-3177-4_6

[9] Rahat IS, Ghosh H, Shaik K, Khasim S, Rajaram G. Unraveling the Heterogeneity of Lower-Grade Gliomas: Deep Learning-Assisted Flair Segmentation and Genomic Analysis of Brain MR Images. EAI Endorsed Trans Perv Health Tech [Internet]. 2023 Sep. 29 [cited 2023 Oct. 2];9. https://doi.org/10.4108/eetpht.9.4016

[10] Ghosh H, Rahat IS, Shaik K, Khasim S, Yesubabu M. Potato Leaf Disease Recognition and Prediction using Convolutional Neural Networks. EAI Endorsed Scal Inf Syst [Internet]. 2023 Sep. 21 https://doi.org/10.4108/eetsis.3937

[11] Mandava, S. R. Vinta, H. Ghosh, and I. S. Rahat, "An All-Inclusive Machine Learning and Deep Learning Method for Forecasting Cardiovascular Disease in Bangladeshi Population", EAI Endorsed Trans Perv Health Tech, vol. 9, Oct. 2023.

https://doi.org/10.4108/eetpht.9.4052
[12] Mandava, M.; Vinta, S. R.; Ghosh, H.; Rahat, I. S. Identification and Categorization of Yellow Rust Infection in Wheat through Deep Learning Techniques. EAI Endorsed Trans IoT 2023, 10. https://doi.org/10.4108/eetiot.4603
[13] Khasim, I. S. Rahat, H. Ghosh, K. Shaik, and S. K. Panda, "Using Deep Learning and Machine Learning: Real-Time Discernment and Diagnostics of Rice-Leaf Diseases in Bangladesh", EAI Endorsed Trans IoT, vol. 10, Dec. 2023 https://doi.org/10.4108/eetiot.4579
[14] Khasim, H. Ghosh, I. S. Rahat, K. Shaik, and M. Yesubabu, "Deciphering Microorganisms through Intelligent Image Recognition: Machine Learning and Deep Learning Approaches, Challenges, and Advancements", EAI Endorsed Trans IoT, vol. 10, Nov. 2023. https://doi.org/10.4108/eetiot.4484
[15] Mohanty, S.N.; Ghosh, H.; Rahat, I.S.; Reddy, C.V.R. Advanced Deep Learning Models for Corn Leaf Disease Classification: A Field Study in Bangladesh. Eng. Proc. 2023, 59, 69. https://doi.org/10.3390/engproc2023059069
[16] Alenezi, F.; Armghan, A.; Mohanty, S.N.; Jhaveri, R.H.; Tiwari, P. Block-Greedy and CNN Based Underwater Image Dehazing for Novel Depth Estimation and Optimal Ambient Light. Water 2021, 13, 3470. https://doi.org/10.3390/w13233470