Expert Systems With Applications 214 (2023) 119128

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Bayesian game model based unsupervised sentiment analysis of product reviews

Neha Punetha, Goonjan Jain *

Department of Applied Mathematics, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India

ARTICLE INFO

Keywords:
Sentiment Analysis
Game Theory
Bayesian Game
Nash Equilibrium
SentiWordNet

ABSTRACT

Sentiment Analysis is a task of computationally recognizing and contextualizing opinions stated in a text. We mainly assess whether the writer's attitude towards a specific topic, or a product, is positive, negative, or neutral. Numerous machine learning and fuzzy logic methods have been reconnoitered for sentiment analysis. Yet, the application of mathematical optimization techniques for sentiment tagging is still unexplored. This study presents a novel mathematical framework for sentiment analysis of reviews based on Game Theory. We identify whether the sentiment of a review is positive or negative. In the first step, we comprehend a review and derive context scores from review comments using the SentiWordNet lexicon. We comprehensively combine the computed context and rating scores using the Bayesian Game Model to deduce the sentiment of reviews. Experimental results on three benchmark review datasets, viz. Food, Mobile, and Electronics demonstrate that the proposed model yields state-of-the-art results. We also statistically validated the stability and correctness of the results. The proposed model ensures rational and consistent results. The utility of the game theory model for sentiment analysis creates a new paradigm for diverse NLP tasks.

1. Introduction

Sentiment Analysis is a Natural Language Processing (NLP) task to identify, extract, and quantify text sentiments. Computerized methods decipher text sentiments or views (Agarwal, Mittal, Bansal, & Garg, 2015). In sentiment analysis, we study a text's subjective information and examine people's feelings, opinions, emotions, or attitudes toward a product (Bu, Li, Cao, Wu, & Zhang, 2016; Hussein, 2018). Sentiments are classified as positive, negative, or neutral. For example,

- "I like the new design of your dress" → Positive Sentiment.
- "I do not like the new design" → Negative Sentiment.
- "The new design is okay!" → Neutral Sentiment.

Online purchasing has grown dramatically over the previous decade due to increased internet usage (Anatomy & Mahadevan, 2000; Ray, 2018). Most online shoppers prefer to examine product reviews before making a purchase decision. Both positive evaluations and reasonable responses boost sales (Contratres, Alves-Souza, Filgueiras, & DeSouza, 2018). Sentiment analysis may give firms helpful information to build efficient business strategies (Dwivedi & Pant, 2019). Firms can use sentiment analysis to monitor brand and product performance, resolve customer complaints, gather in-depth information for strategic analysis[1], and enhance product quality (Atoum, 2020). Sentiment Analysis can help discover selling possibilities, reduce churn, boost client acquisition, increase customer retention, and resolve customer complaints. Thus, sentiment classification is a crucial task (Cao, Xu, Yin, & Pan, 2022).

Sentiment analysis can be performed in a subjective or objective context. Subjective context gives better insight into the author's sentiment than objective context. An objective context or perspective demonstrates ordinary statements or facts without emotions, feelings, or mood (Jeon, Xia, & Liu, 2011; Sarkar, 2019). Subjective context exhibit a particular mood, emotion, or feeling (Mee, Homapour, Chiclana, & Engel, 2021). There are varieties of techniques for performing sentiment analysis of subjective contexts. In the proposed model, we classify the subjective context of reviews as positive or negative.

The vast content published daily on several social networking websites necessitates automated solutions for managing and identifying all opinions. Manually analyzing the sentiments can be highly tedious and

* Corresponding author.
E-mail address: goonjanjain@dtu.ac.in (G. Jain).

[1] https://www.repustate.com/blog/sentiment-analysis-real-world-examples/.

https://doi.org/10.1016/j.eswa.2022.119128
Received 16 March 2022; Received in revised form 21 October 2022; Accepted 21 October 2022
Available online 27 October 2022
0957-4174/

sometimes impossible due to the sheer volume of data (Jain, Pamula, & Srivastava, 2021; Jain, Suvarna, & Jain, 2021). As a result, various sentiment analysis algorithms and techniques have been developed. In the proposed model, we introduce an unsupervised Bayesian Game model to perform sentiment analysis of product reviews. In this study, we create a novel decision-making system that labels each review with a polarity sentiment. This study employs a mathematical optimization technique to perform sentiment analysis of reviews. The Game Theory-based Bayesian approach is used to maximize the accuracy and efficiency of the sentiment classification task.

Game Theory is a framework for conceptualizing social situations, including competing players, and determining the best decision-making of autonomous and competing agents in a strategic context. Game Theory is the study of the formulation of optimal strategies in conflict. A Bayesian Model in Game Theory is a game that combines characteristics of Bayesian probability to simulate the result of player interactions. Game Theory and Bayesian game are explained in subsections 3.1 and 3.2, respectively.

1.1. Research gap

Previous researches demonstrate supervised, semi-supervised and unsupervised approaches for sentiment analysis. However, they have their limitations, as given below.

Machine learning techniques offer superior classification precision and accuracy. However, they also have significant limitations, such as they are time-consuming, domain-dependent, and requiring human intervention and information labeling. SVM and ANN-based methods are also explored for sentiment classification in the literature. However, SVM performs poorly when there are more characteristics than samples and fails to work when there is no probabilistic reason for the classification. On the contrary, ANN requires a lot of training time and a massive dataset (Aldoğan & Yaslan, 2017; Berka, 2020; Fiok, Karwowski, Gutierrez, & Wilamowski, 2021; Ghiassi, Skinner, & Zimbra, 2013).

Deep learning models enhance the precision of sentiment analysis, but as the number of hidden layers increases, the amount of processing and storage requirements also increases. Deep learning techniques require a large corpus and efficient word embedding. These models take longer to train because of their sophisticated design, high computational cost, and overfitting issues (Basiri, Nemati, Abdar, Cambria, & Acharya, 2021; Birjali, Kasri, & Beni-Hssane, 2021; Hossain, Bhuiyan, Tumpa, & Hossain, 2020; Li, Chen, Zhao, & Li, 2021).

Semi-supervised methods like generative, co-training, and graph-based methods are also used for sentiment analysis. However, classification problems do not work well with generative approaches. Co-training is not suitable for datasets with only one feature sensitive to noise. Graph-based approaches depend on the structure and weights of edges and do not work well if the graph does not fit the task. Hybrid methods require less time but lack reliability and precision (Birjali et al., 2021; Jindal & Aron, 2021; Lin, He, Everson, & Rüger, 2012; Zhang & Singh, 2014).

Unsupervised approaches for sentiment analysis include hierarchical and partition methods. Due to the high computation cost, hierarchical algorithms are not ideal for massive datasets. Partition methods are relatively scalable and straightforward but have poor accuracy and stability and are sensitive to noisy data. The lexicon-based approach does not need information labeling but depends upon the domain, needs powerful linguistic resources, and has low precision (Birjali et al., 2021; García-Pablos, Cuadros, & Rigau, 2017; Jindal & Aron, 2021; Kim, Zhang, Chen, Oh, & Liu, 2013; Lin et al., 2012). Many unsupervised techniques include fuzzy-based methods to classify the sentiments of English text. The primary disadvantage of fuzzy logic is its reliance on human intelligence and expertise. Due to the unreliable and varying fuzzy membership values, fuzzy-based approaches are not dependable, and their results vary with changes in fuzzy values (Chiha, Ayed, & da Pereira, 2022; Jain & Lobiyal, 2022; Karthik & Ganapathy, 2021; Vashishtha and Susan, 2019a,b). Various studies used different models. But none has applied mathematical optimization techniques in sentiment tagging.

1.2. How the proposed algorithm is different from others?

We designed an unsupervised mathematical optimization algorithm to assign sentiment tags to reviews in the proposed work. We comprehend the implicit information of textual feedback of reviews and calculate its context scores. We use ratings and context scores of reviews to perform sentiment analysis. The sentiment tag of each review is deduced using the principle of Bayesian Nash Equilibrium (BNE). The Proposed Bayesian Model has a better generalization than the other approaches proposed in the literature. The proposed model is language-independent and can be implemented for other languages with minimal alterations. The condition is that we have SentiWordNet (Esuli & Sebastiani, 2006) or a similar knowledge base to compute the context scores of a review comment in that language. Also, the proposed model is domain-independent and can be applied to any domain dataset having review comments and ratings. The results demonstrated state-of-art results over three English and two general (English and Hindi) datasets. This study introduced the application of mathematical optimization for sentiment tagging of the reviews.

1.3. Contributions

The contributions and novelties of the proposed study are as follows:

1) To the best of our knowledge, this is the first study to perform sentiment classification of reviews with the help of a Game Theory model. This research assesses the suitability of game models for NLP applications like sentiment classification of English text.
2) We propose a sentiment tagger that assigns the appropriate sentiment tag to a review. We use a review's star rating and textual feedback to perform Binary class sentiment classification (positive and negative) of review datasets.
3) The proposed model maximizes accuracy while minimizing time and space complexity. Since the proposed model is unsupervised, it is domain and language-independent and can be applied to any low-resource language dataset with minimal adjustments.
4) We also examine and establish the statistical significance of the proposed framework. Evaluation performance indicators are studied and applied to the proposed model to determine its effectiveness and provide state-of-the-art performance.

1.4. Organization

The paper is structured as follows: The fundamentals of sentiment analysis, research gaps, and contributions are presented in *Section 1*. *Section 2* provides the latest research in sentiment analysis and game theory. *Section 3* discusses the proposed Bayesian model for sentiment analysis. *Section 4* evaluates the proposed model's performance on three review datasets and compares it to several state-of-the-art methodologies. The error analysis, computational complexity, and challenges of the proposed model are discussed in *Section 5*. *Section 6* finishes with some insightful findings and recommendations for the future.

2. Related work

This section discusses prior sentiment analysis work and the application of game theory in the field of NLP in chronological order.

Research in sentiment analysis has gained much interest in the previous decade. Ravi et al. (2015) developed many applications of sentiment mining, like lexicon construction and opinion word extraction. Fang and Zhan (2015) performed sentence-level and review-level categorizations with promising outcomes. García-Pablos et al. (2017)

introduced W2VLDA, an almost unsupervised system based on topic modeling. Donadi (2018) presented a sentiment analysis system for the German language utilizing TensorFlow. Vashishtha and Susan (2019a,b) proposed a fuzzy system that used a novel unsupervised nine fuzzy rule-based system to classify the text into positive, negative, or neutral sentiment classes, combining NLP techniques and Word Sense Disambiguation. Da'U and Salim (2019) introduced a sentiment-aware deep recommender system with a neural attention network (SDRA) to improve recommendation system performance. Song, Park, Shin, and Shik (2019) used sentiment lexicon embedding to develop an attention-based extended short-term memory network for aspect-level sentiment analysis. Li et al. (2019) combined textual information with sentiment time series to achieve multi-document sentiment prediction.

Yang, Ma, Zhang, Gao, and Xu (2020) proposed a novel model named Attention-based Point Network for sentiment analysis. Rani and Lobiyal (2020) discovered sentiment polarity and recommended statistical and knowledge-based approaches to build corpus-specific stop word lists for Hindi texts. Usama et al. (2020) proposed a new model based on RNN with a CNN-based attention mechanism for sentiment analysis. Pandesenda et al. (2020) developed a novel attention-based model to solve the problem of how to use NLP for online reviews. Akhtar, Ekbal, and Cambria (2020) used a multi-layer perceptron network and proposed a stacked ensemble method for predicting the degree of intensity of emotion and sentiment. Liu and Shen (2020) implemented the Recurrent Memory Neural Network (ReMemNN), evaluated a wide variety of datasets on three English and four Chinese datasets from different sources, and the approach outperformed the state-of-art performance. Basiri et al. (2021) proposed a bidirectional Neural Network architecture for sentiment analysis to tackle the disadvantages of LSTMs and Gated Recurrent Units (GRUs). Zhao and Yu (2021) developed a knowledge-enabled BERT model for aspect-based sentiment analysis. Mee et al. (2021) presented a methodology for assigning a political value to an MP based on their voting record on a single issue (Brexit). They used regression and sentiment analysis, precisely Term Frequency–Inverse Document Frequency (TF–IDF). The study by Dahooie et al. (2021) used an integrated framework that combined sentiment analysis and multi-criteria decision-making techniques. Perikos, Kardakis, and Hatzilygeroudis (2021) presented attention-based models based on (RNNs) and investigated their performance in various sentiment analysis contexts. To incorporate relations between dimensions into deep neural networks for dimension score prediction, Xie, Lin, Lin, Wang, and Yu (2021) proposed a multi-dimensional relation model.

Ileri and Turan (2021) employed a neural network for sentiment analysis. Nguyen, Phan, and Do (2021) proposed embedding information in the ontology to represent the main aspects of the dataset in the word embedding layer of sentiment classification deep learning algorithms. On the ABSA task, Wu, Ma, Chiclana, Liu, and Wu (2022), Wu, Zhang, Shi, Wu, and Song (2022) introduced the PD-RGAT model, a relational graph attention network built on the phrase dependency graph aggregating directed dependency edges and phrase information. Carosia et al. (2021) implemented investment strategies on the Brazilian stock market to predict sentiment analysis using deep learning. A study by Vashishtha and Susan (2021) proposed unsupervised sentiment analysis using SentiWordNet and fuzzy entropy. Unsupervised sentiment analysis methods were proposed for Twitter accounts (Fiok et al., 2021). Latent Dirichlet Allocation (LDA) methods were used in aspect-based sentiment analysis by Ozyurt and Akcayol (2021). Liang, Su, Gui, Cambria, and Xu (2022) developed a convolutional graph network based on SenticNet that leverage the affective dependencies of the sentence. Swathi, Kasiviswanath, and Rao (2022) presented TLBO and LSTM models for stock price prediction using the Twitter dataset. Wu and Chung (2022) proposed the DVA-BERT model for sentiment analysis which outperforms the BERT model. Wu, Ma et al. (2022), Wu, Zhang et al. (2022) introduced the OVO-SVM algorithm and Word2Vec for hotel selection. Li, Johnson, Aarhus, and Shah (2022) conducted sentiment analysis on MOOCs platforms.

In recent years, we have seen a growth in the application of Game Theory in the field of NLP. Bu et al. (2016) proposed an effective Game Theory-based emotional evolution prediction system. Tripodi et al. (2016) and Jain and Lobiyal (2022) presented the application of Game theory for word sense disambiguation. Ahmad and Ahmad (2019) created a multi-document summarization framework based on Game Theory integrated with Wikipedia. Liu and Li (2019), Jain, Katarya, and Sachdeva (2020), Huang, Yang, Li, Yang, and Tang (2021), and Xiao et al. (2022) introduced various methods for rumor detection using Game theory. Ruseti et al. (2020) and Jain, Gayathri, and Ranjan (2022) proposed game theory-based sentiment analysis models. Jain et al. (2020) proposed a GOLD algorithm to identify opinion leaders in online social networks. Saxena, Mangal, and Jain (2021) performed keyword extraction using Game Theory. A study published by Jain, Pamula et al. (2021); Jain, Suvarna et al. (2021) improved accuracy and efficiency for query expansion using evolutionary Game Theory. Barfar (2022) used Shapley value approaches to detect propaganda.

Various studies used the Bayesian game model for decision-making in various fields and had good outcomes. Sharma and Spaan (2012) proposed a Bayesian game-based model for decision-making in a stochastic environment. Dahiya and Gupta (2021) proposed a Bayesian game mechanism-based pricing scheme to impose a penalty. Xia et al. (2020) applied the Bayesian model for Vehicle-to-Vehicle Electricity Trading. Abapour, Mohammadi-Ivatloo, and Tarafdar Hagh (2020) used Bayesian game strategies for bidding in the electricity market and used the Bayesian mechanism for cyber security in industries.

The studies motivated us to explore game theory's utility, specifically the Bayesian model, for NLP tasks. This paper introduces an innovative unsupervised method for sentiment analysis. We use the Bayesian game to evaluate text sentiment by integrating context and the rating of written reviews.

3. Proposed methodology

This section introduces Game Theory in subsection 3.1. The proposed Bayesian Game Model for sentiment analysis is given in subsection 3.2. Subsection 3.3 contains the proposed methodology for sentiment analysis. We give a detailed illustrative example of the proposed model in subsection 3.4.

3.1. Game Theory

von Neumann and Morgenstern (2007) introduced the Game Theory to build a mathematical framework capable of simulating decision-making fundamentals in interactive scenarios (Muthumanickam & Ilavarasan, 2015; Vincent & Brown, 2005). Game Theory is described as a set of situations in which each player must consider the actions of other players to make sensible judgments (Biaou et al., 2020). All players are assumed to be rational. The action is determined by other players' actions (Vincent & Brown, 2005). In interactive decision situations, game theory provides predictive power. We can create and evaluate complicated interactions involving decision-makers[2], such as financial market investors, supply-chain businesses, and governments with competing interests (Arsenyan, Büyüközkan, & Feyzioħlu, 2015; Vincent & Brown, 2005). A game is represented as $\Gamma = <N, (S_i), (u_i)>$, where, N = Number of players, (S_i) = Set of strategies, and (u_i) = Payoffs. A game has the following components:

- **Players:** Participants in a game who aim to maximize their payoffs.
- **Strategy profile:** A set of players' actions is known as a strategy profile. The game's strategy profiles of the player are displayed simultaneously by $m^* = \{m_1^*, m_2^*, \cdots, m_i^*\}$ the set of action sets for i^{th}

[2] https://www.scientificamerican.com/article/what-is-game-theory-and-w/.

player plays in the game. The strategy profile of opponent players is $m^*_{-1} = \{m^*_1, m^*_2, \cdots m^*_{i-1}, m^*_i\}$.

- **Payoff:** The payout of player receives after arriving at a particular outcome. The payout can be in any quantifiable form.
- **Outcome:** The participants determine an ideal approach for each action.
- **Best Response:** Each player's best response (*br*) is the best reaction he makes to a specific action given by Eq. (1).

$$br_i(m_{-i}) = \{m_i \in A_i : u_i(m_i, m_{-i}) \geq u_i(m_i^{\ k}, m_{-i}); \forall m_i^{\ k} \in A_i\} \tag{1}$$

- **Nash Equilibrium:** The Nash equilibrium is a decision-making principle that states players can achieve the desired outcome by not deviating from their initial strategy. It is one of the game's best responses. i.e., if $m*$ is a Nash equilibrium $\Rightarrow m^*_i \in br_i(m^*_{-i})$. Alternatively, Nash equilibrium implies the intersection of the Best responses of players. Along with the above components, some strategies are essential to solve any game under Game Theory.

Strongly Dominated Strategy: A dominant strategy is a strategy that consistently generates the best outcome for the player regardless of what other players are doing. Every strongly dominant strategy implies Nash equilibrium and thus the best response. A strategy s_i, $s_i \in S_i$, is said to be strongly dominated if there exists another strategy whose payoff value is highest among the rest strategies.

$$u_i(s_i^*, s_{-i}) > u_i(s_i, s_{-i}) \forall s_{-i} \in S_{-i} \text{ for some } s_{-i} \in S_{-i} \tag{2}$$

In such a case, we say the strategy s_i^* strongly dominates strategy S_i given by Eq. (2) where S_i = Set of actions of the i^{th} player, S_{-i} = Set of actions of any other player.

Weakly Dominant Strategy: A strategy s_i, $s_i \in S_i$, is said to be weakly dominated by a strategy $s^*_i \in S_i$, if.

$$u_i(s_i^*, s_{-i}) \geq u_i(s_i, s_{-i}) \forall s_{-i} \in S_{-i} \textit{ and } u_i(s_i^*, s_{-i}) > u_i(s_i, s_{-i}) \textit{ for some } s_{-i} \in S_{-i} \tag{3}$$

The strategy s_i^* is said to be a weakly dominant strategy, as illustrated in Eq. (3).

Example: Prisoner's Dilemma Problem: Two suspects are apprehended for a crime and are in separate rooms in a police station, with no means of communicating. The prosecutor has separately told that they have two options either to confess (*C*) or not to confess (*NC*). A matrix represents the game. A row represents strategies for the first player. The column represents the strategies of the second player. Matrix entry, (*x,y*) where *x* is the payoff of the first player and *y* is the payoff of the second player.

Let *N* be the no. of players, with each player having a set of strategies. In the given example, the number of players is two, $N = 2$. Each player has 2 strategies confess or not confess $S_1 = S_2 = \{C, NC\}$. Each cell represents the payoffs of players for choosing a specific strategy. For example, if both players choose *NC, NC,* their respective payoffs are (−2, −2). The complete payoff matrix is shown in Table 1.

Strategy *C* is a strongly dominating strategy for *player 1* because the payoff values (u_i) of the *C* strategy are more significant than the *NC* strategy.

Table 1
Matrix form representation of Prisoner's Dilemma.

		Player 2	
		NC	C
Player 1	**NC**	−2, −2	−10, −1
	C	−1, −10	−5, −5

$u_1(C, NC) > u_1(NC, NC)$ and $u_1(C, C) > u_1(NC, C)$

Similarly, for *player 2*, strategy *C* strongly dominates strategy since the payoff values of the *C* strategy are more significant than the *NC* strategy.

$u_2(NC, C) > u_2(NC, NC)$ and $u_2(C, C) > u_2(C, NC)$

Thus, *C* is the strongly dominant strategy for both *player 1* and *player 2*. Thus (*C, C*) is a strongly dominant strategy. It is also the Nash equilibrium and the best response of the players in this prisoner's dilemma problem. The Game Theory encompasses a wide range of models. In the proposed work, we used the Bayesian Game Theory model to analyze the sentiments of reviews.

3.2. Proposed Bayesian model

In a Bayesian game, a player may not know the exact payoff functions of the other players but has ideas about each other's payoff functions. Players can only guess the strategies of their opponents. The strategies are limited to a general understanding of the probability distribution of the opponent's strategy profiles. The Bayesian Game Model includes type spaces, action spaces, payoff functions, and prior beliefs. A player's strategy is a comprehensive plan of action that addresses every possibility that may emerge for the player (Narahari, Narayanam, Garg, & Prakash, 2009). In a Bayesian game, it is unknown which player will select which type of strategy; hence, a prior probability distribution is used to predict other players' strategies. To achieve Nash equilibrium in Bayesian games, BNE is used (Narahari et al., 2009). BNE is an extended variation of Nash equilibrium designed to achieve equilibrium under uncertainty. Each player must create an optimum gaming strategy to maximize their utility against other players' randomized mixed strategies.

We propose a Bayesian Game Model to perform sentiment analysis of review comments. The Bayesian Game Model can be mathematically represented by a tuple $G = <N, (T_i), (A_i), (p_i), (u_i)>$. The main components of the proposed Bayesian model are given below:

- The number of reviews is *N (1,2,3,4, …n)*, and each review is treated as a player.
- T_i denotes the type of ways sentiment of a review can be evaluated $T_i \in \{C, R\}$ for $i = 1,2,3 \ldots\ldots\ldots n$, where '*C*' denotes the context and '*R*' denotes the rating.
- A_i is a collection of actions performed to achieve a specific goal. Actions can be either positive or negative, i.e., $A_i \in \{P_i, N_i\}$.
- p_i is the probability of occurrence of the type of review T_i and is given by Eq. (4) and Eq. (5). The probability of occurrence of rating type is

$$P(R) = p \tag{4}$$

and context type is.

$$P(C) = 1 - P(R) \Rightarrow P(C) = 1 - p \tag{5}$$

If $p = 1$, then $1 - p = 0$. In such a case, we smooth the value of probabilities by re-evaluating $p = 0.9$ and $1 - p = 0.1$.

- The utility function u_i: $T_i \times A \rightarrow \mathbb{R}$ indicates the payoff of each review for any action profile or type. $A = A_1 \times A_2$. Our algorithm has the same set of actions for both the context and rating of a review. However, the review's probability of taking various actions might vary depending on the context and rating. We consider that the context of a review is of two types: positive context (C_P) and negative context (C_N), and the rating of a review is also of two types: positive rating (R_P) and negative rating (R_N).
- The final sentiment tag of the review is deduced by BNE, which is each review's best response.

Definition 1. ***Bayesian Nash Equilibrium***: *BNE is described as a strategy profile that optimizes each player's expected payoff given their beliefs and the strategies used by the other players. For games of incomplete information, BNE is an extension of Nash equilibrium. A strategy of the i^{th} player is a function $s_i : T_i \rightarrow A_i$. s^* is a BNE of the Bayesian game $G = <N, (T_i), (A_i), (p_i), (u_i)>$ if and only if for each $i \in N$ and each $t_i \in T_i$,*

$$U(s^*; t_i) \geq U(a_i, s^*_{-i};\ t_i) \quad \forall a_i \in A_i \tag{6}$$

$$\text{Where } U_{t_i}(s_{t_i}, s_{-t_i}) = \phi_{t_{-i}}[U_i(s_i(t_i), s_{-i}(t_{-i}); t_i, t_{-i})] \tag{7}$$

$$U_{t_i}(s_{t_i},\ s_{-t_i}) = \sum_{t_{-i} \in T_{-i}} p_{probabilty_{(i)}}(t_{-i} | t_i).u_i(s_1(t_1), (s_2(t_2) \ldots, s_n(t_n); t_i, t_{-i}) \tag{8}$$

where U_{t_i} refers to the utility/expected payoff of player i conditioned on the type t_i given by Eq. (8).

3.3. Proposed algorithm for sentiment analysis

To perform sentiment analysis, we use two parameters of a review: textual comments and customer ratings. The rating is already in numeric form. We convert the textual comments to a numeric value called context score. Context scores are numerical values giving the review's polarity score for positive and negative sentiments. We use SentiWordNet to calculate the context scores. Then, we feed the rating and context scores to the proposed model that predicts the sentiment of reviews. Fig. 1. shows the flowchart depicting the steps to implement the proposed model. Table 2. encapsulates the notations used in this study.

Table 2
Notations used in this study.

Notations	Description
$(A_j)_P$	Positive Action performed by R_j
$(A_j)_N$	Negative Action performed by R_j
$(A_i)_{C_P}$	Context type of review R_i while performing positive action.
$(A_i)_{C_N}$	Context type of review R_i while performing negative action.
$(A_i)_{R_P}$	Rating type of review R_i while performing positive action.
$(A_i)_{R_N}$	Rating type of review R_i while performing negative action.
C_P	Context is positive
C_N	Context is negative
R_P	Rating is positive
R_N	Rating is negative

Step 1: Determine context scores of review comments

In this step, we determine the context score of textual comments using SentiWordNet. SentiWordNet contains a list of words with respective positive and negative polarities values. Fig. 2. shows an excerpt of SentiWordNet.

We take content POS-tagged words of a review comment. We extract and add each word's positive and negative polarity values. Further, we compute the mean of the positive and negative polarities to get the context scores of a review. We follow Algorithm 1 to generate context

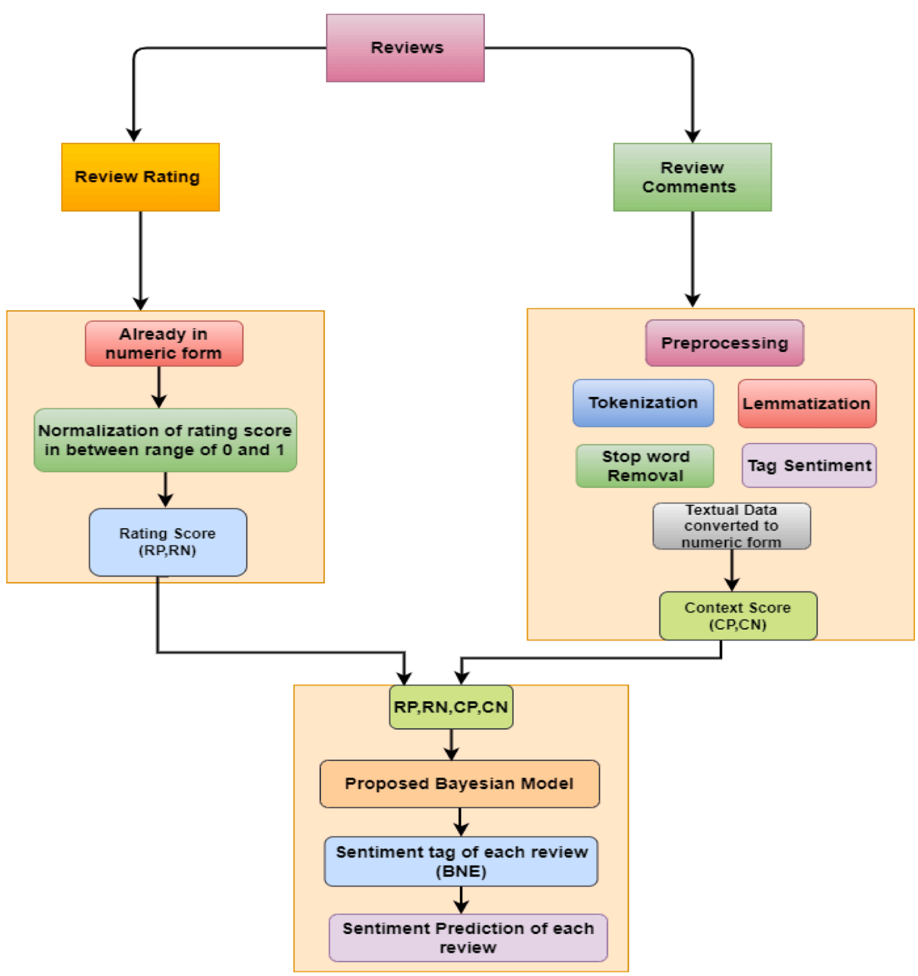

Fig. 1. Flowchart of the proposed model for sentiment analysis of reviews.

POS	ID	PosScore	NegScore	SynsetTerms	Gloss
a	00005107	0.5	0	uncut#7	complete; "the full-length play"
a	00005718	0.125	0	infinite#4	total and all-embracing;
a	00006245	0	0	relational#1	having a relation or being related
a	00016532	0	0.375	torrential#3	pouring in abundance; "torrential rains"
a	00016647	0.125	0.5	verdant#1	characterized by abundance of verdure
a	00017688	0.375	0.25	unabused#1	not physically abused; treated properly
a	00016247	0.125	0.5	superabundant#1	most excessively abundant
a	00018069	0.25	0	bankable#2	acceptable to or at a bank;
a	00020103	0.125	0	remote#4 outback#1	inaccessible and sparsely populated;
a	00013887	0	0.25	abundant#1	present in great quantity;

Fig. 2. An excerpt of SentiWordNet lexicon.

scores.

Algorithm 1: Calculate context scores of reviews.

Input: W - a set of open class and parts of speech-tagged words extracted from a review, SWL – SentiWordNet list of words with positive and negative sentiment value.
Output: Context Scores of i^{th} review $C_i = \{C_P, C_N\}$, where C_P = positive context score and C_N = negative context score.
1: Initialize, $C_P = 0$ and $C_N = 0$; such that $C_i = \{0, 0\}$.
2: Take $W = (w_1, w_2, \ldots, w_i, \ldots, w_n)$ where w_i represents the i^{th} $(1 \leq i \leq K)$ word in the input review.
3: Repeat step 4 for each word of W
4: If ($w_i \in SWL$), then
$P_{Sentiscore} = p +$ *positive sentiment score of* w_i
$N_{Sentiscore} = n +$ *negative sentiment score of* w_i
5: C_p and C_N are calculated by averaging $P_{Sentiscore}$ and $N_{Sentiscore}$.
$C_P = \frac{P_{Sentiscore}}{K}$ and $C_N = \frac{N_{Sentiscore}}{K}$ // K is the number of words in set W

Step 2: Normalize context and rating scores

The context score ranges between 0 and 1, and the rating has values between 1 and 5. To reduce the dominance of one parameter over the other, we normalize the scores by dividing each value of rating and context by the maximum numeric value of rating and context in the dataset. Thus, the normalized context and rating score values are between 0 and 1. The values of C_P and C_N are evaluated using Algorithm 1. Value of R_P = Given a rating of a review and $R_N = (5 - R_P)$. We normalize C_P, C_N, R_P, and R_N scores such that their values range from 0 to 1.

Step 3: Play Bayesian games among reviews

Further, the Bayesian game is played between two players (reviews), R_i and R_j. We consider R_i has two types, either "context type" or "rating type." The R_j has only one type. The R_i knows its type and the R_j*'s* type, but the R_j does not know the R_i*'s* type. Thus, R_j has incomplete information about R_i*'s* type and action, making it a Bayesian game. R_j just knows the probability of R_i playing type and action. The context and rating are two independent events, and the sum of the probability of occurrence of all independent events is 1. So, if p is the probability of occurrence of R_i for performing rating type, then $(1\text{-}p)$ is the probability of R_i for performing context type. Both players R_i and R_j are aware of this probability. That is, if the rating of a review is *4*, then the probability of occurrence of rating type $P(R) = 4/5 = 0.8$. Thus, the probability of occurrence of context type is $P(C) = 1\text{–}0.8 = 0.2$. In Fig. 3., α and ψ represent numeric values for the positive context of R_i and R_j. γ and ω denote the numeric values for the negative context of R_i and R_j. Similarly, β and Ω depict the numeric value of the positive rating of R_i and R_j. And δ, and $\in$ denote the numeric value of the negative rating.

To play the Bayesian game, we compute the values of the four matrices, as shown in Fig. 4. In each matrix, a row represents an action of the first player, and a column represents an action of the second player. Each matrix entry is (x,y), where x is the payoff of the first player and y is the payoff of the second player. The first matrix is the context type matrix. In Context Matrix, we denote the context type of a review R_i with (T_{ic}). The review (R_i) can perform two actions, either positive $(A_i)_{C_P}$ or negative $(A_i)_{C_N}$. The second matrix is the rating type matrix. The Rating matrix denotes the rating type of review R_i with (T_{iR}), where R_i can perform either positive $(A_i)_{R_P}$ or negative $(A_i)_{R_N}$ actions. The review R_j can take either positive $(A_j)_P$ or negative $(A_j)_N$ actions.

We assign the probability of occurrence of Context Matrix and Rating Matrix as $P(T_{ic}) = 1 - p$ and $P(T_{jR}) = p$, respectively. Since the context and rating are mutually exclusive so the sum of the probabilities of the context and rating matrix should be one. Thus, the combined probability of the Context matrix and Rating matrix is $P(T_{jR}) + P(T_{jc}) = 1$. The third and fourth matrices are derived by multiplication of each entry of type matrices with their respective probabilities.

The combined context matrix and rating matrix are shown in Fig. 5. The combined matrix's entries are the payoff values for the game. Payoffs are averaged as per the probabilities of each type of review. We compute the payoffs using the formulas mentioned in each cell of the combined matrix. There are two formulas in each cell, the first formula evaluates the payoff of the first review, and the second formula evaluates the payoff of the second review.

Step 4: Deduce the sentiment tag of a review

After getting payoffs from step 3, we use Eq. (2) and Eq. (3) to find the players' dominant strategies. We find the BNE and deduce the sentiment tag for the reviews. We follow Algorithm 2 to deduce the sentiment tag of a review.

Algorithm 2: Deduce sentiment tag (positive or negative) for review.

Input: Reviews data file R with context score $\{C_P, C_N\}$ and the rating score $\{R_P, R_N\}$.
Output: Tagged Sentiment, i.e. $\{R_i, R_j\} \in \{P, N\}$, where a set of strategies $\{s_{-i}, s_i\} \in \{P, N\}$.
1: Make two matrices of context $\{C_P, C_N\}$ and rating $\{R_P, R_N\}$ with strategy positive & negative.
2: Calculate payoffs of R_i and R_j, $U_i(R_i, R_j) = (1\text{-}p)u_i(R_i, R_{jC}) + (p)\, u_i(R_i, R_{jR})$, where p is probability.
3: Calculate the combined matrix of context and rating by multiplying with respective probabilities.
4: Apply Game Theory strategies and evaluate the best response.
If $U_{t_i}(s_i^*(t_i), s_{-i}(t_{-i})) > U_{t_i}(s_i, s_{-i}(t_{-i})) \,||\, U_{t_i}(s_i^*(t_i), s_{-i}(t_{-i})) \geqslant U_{t_i}(s_i, s_{-i}(t_{-i}))$
Then $U_{t_i}(s_{t_i}, s_{-t_i}) = \sum_{v_{-i} \in T_{-i}} p_{robability(i)}(v_{-i}|t_i) U_i(t_i, v_{-i}, s_{t_i}, s_{v_{-i}})$
5: The best response strategy is the sentiment tag of individual review. If $s_i(t_i) \rightarrow BNE$, then
$s_i(t_i) \in \underset{s_i' \in S_i}{\operatorname{argmax}} \sum_{t_{-i}} p_{robabilty(i)}(t_{-i}|t_i) U_i(s_i', s(t_{-i}), t_i, t_{-i})$
6: Repeat steps 1 to 5 for all combinations of reviews in the data file.

3.4. Illustrative example

Let us consider two reviews, R_1 and R_2.

R_1 (4 stars): "*I have buy several of the Vitality can dog food product and have find them all to be of good quality. The product looks more like a stew than a processed meat and it smell better. My Labrador be finicky and she appreciate this product good than most.*"
R_2 (1 star): "*Product arrived labeled as Jumbo Salted Peanuts…the peanuts were actually small sized unsalted. Not sure if this was an error or if the vendor intended to represent the product as Jumbo*".

First, we compute the context scores of each review using SentiWordNet by following Algorithm 1. Table 3. shows the normalized context and rating values of reviews R_1 and R_2 evaluated following step 2. The reviews R_1 and R_2 can interact in two possible ways.

Types → Actions ↓	R_i		Types → Actions ↓	R_j	
	Context Type	Rating Type		Context Type	Rating Type
Positive	α	β	Positive	ψ	Ω
Negative	γ	δ	Negative	ω	$\in$

Fig. 3. Different values of context and rating type and individual actions performed by reviews.

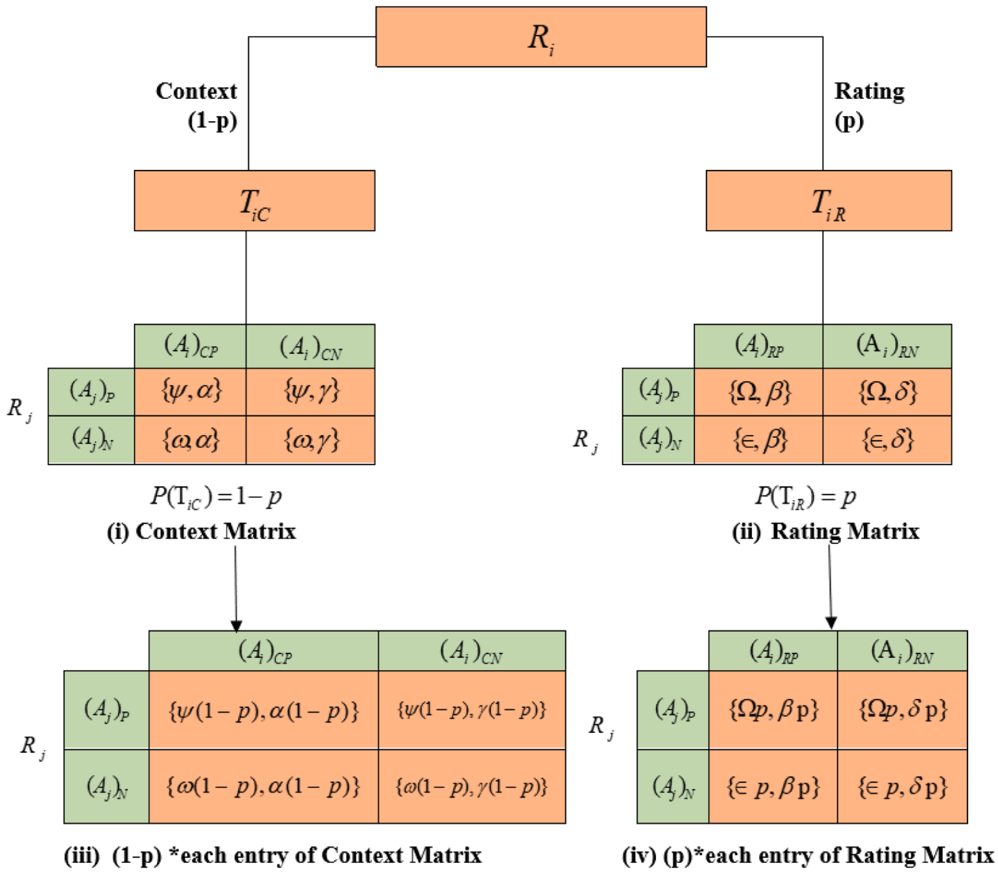

Fig. 4. Demonstration of proposed Bayesian Model.

R_j \ R_i	$(A_i)_{CP}.(A_i)_{RP}$	$(A_i)_{CP}.(A_i)_{RN}$	$(A_i)_{CN}.(A_i)_{RP}$	$(A_i)_{CN}.(A_i)_{RN}$
$(A_j)_P$	$\{(\psi(1-p)+\Omega p), (\alpha(1-p)+\beta p)\}$	$\{(\psi(1-p)+\Omega p), (\alpha(1-p)+\delta p)\}$	$\{(\psi(1-p)+\Omega p), (\gamma(1-p)+\beta p)\}$	$\{(\psi(1-p)+\Omega p), (\gamma(1-p)+\delta p)\}$
$(A_j)_N$	$\{(\omega(1-p)+\in p), (\alpha(1-p)+\beta p)\}$	$\{(\omega(1-p)+\in p), (\alpha(1-p)+\delta p)\}$	$\{(\omega(1-p)+\in p), (\gamma(1-p)+\beta p)\}$	$\{(\omega(1-p)+\in p), (\gamma(1-p)+\delta p)\}$

Fig. 5. Combined Matrix of context and rating type multiplied with respective probabilities.

Table 3
Normalized values of Context and Rating of R_1 and R_2.

Type → Action ↓	R_1		R_2	
	Context	Rating	Context	Rating
Positive	0.1079	0.8000	0.0515	0.2000
Negative	0.0114	0.2000	0.1397	0.8000

Interaction 1: In the first interaction, we consider that R_2 has two types, i.e., context and rating. R_1 and R_2 have common actions (positive and negative) to perform. The interaction between the players R_1 and R_2 is shown in Fig. 6. Player R_2 knows the action and type of R_1, but R_1 does not know the actions and type of R_2, or we can say that R_1 has incomplete information about R_2. This incomplete information is Bayesian in nature. R_1 knows the probability of occurrence of the context and rating type matrices of R_2.

Table 4. and Table 5. show the context type matrix and the rating type matrix with normalized values in each cell for R_1 and R_2, respectively. The first value in each cell is the payoff of R_1, and the second

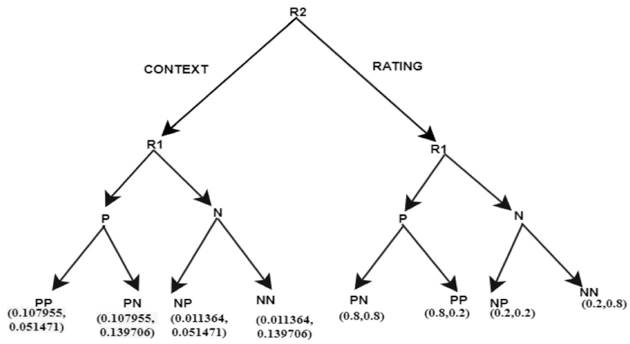

Fig. 6. Extensive form of R_2 and R_1 interaction.

Table 4
Context Type Matrix.

R_2			
R_1		***P***	***N***
	P	(0.1079,0.0515)	(0.1079, 0.1397)
	N	(0.0114,0.0515)	(0.0114,0.1397)

value is the payoff of R_2 while choosing a specific action (*P, N*).

Table 6. is the combined matrix. Payoff values are calculated by multiplying the payoffs in context and rating matrices with respective probabilities of different types (*Section 3.2* from *step 3)*. Next, we apply Eq. (2). and Eq. (3). to evaluate the dominant action of R_1 and R_2 by considering the payoff values that dominate other payoffs. The dominant payoff is **(1.023,0.2718)**, where **1.023** is the payoff value of R_1 and **0.2718** is the payoff of R_2.

Table 7. shows the BNE of interaction 1. The actions corresponding to the BNE are (*P*, *NN*), where *P* is the sentiment tag for R_1 and *NN* is the sentiment tag for R_2. The first *N* in *NN* implies that context chooses negative action, and the second *N* tells that the rating type also chooses the negative action. The interpretation of BNE is displayed in Fig. 7.

Interaction 2: In the second interaction, we consider that R_1 has two types, i.e., context and rating, and both R_1 and R_2 have common actions (positive and negative) to perform. In this case, we reverse the sequence of interactions as illustrated in Fig. 8. and then apply the Proposed Bayesian Model to obtain the BNE.

Table 8. and Table 9. show the Context and Rating Type matrices. Table 10. shows the combined context and rating matrix. Table 11. gives the strongly dominant action of interaction 2, i.e., (*N*, *PP*) where *N* is the best response of R_2 and *PP* is the best response of R_1; first, *P* in *PP* implies context is positive, whereas the second *P* gives that the rating is positive too. Fig. 9. shows the interpretation of BNE.

Table 5
Rating Type Matrix.

R_2			
R_1		***P***	***N***
	P	(0.8000,0.2000)	(0.8000,0.8000)
	N	(0.2000,0.2000)	(0.2000,0.8000)

Table 7
Bayesian Nash Equilibrium: Best response of the individual player.

		R₂
		NN
R₁	***P***	**(1.0231,0.2718)**

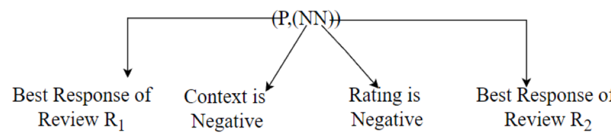

Fig. 7. Interpretation of BNE.

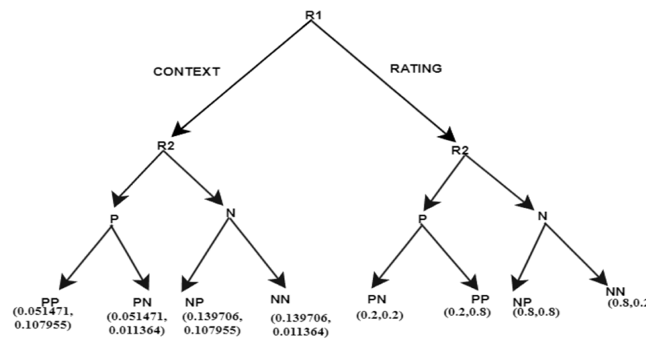

Fig. 8. Extensive form of R_1 and R_2 interaction.

Table 8
Context Type Matrix.

R_1			
R_2		***P***	***N***
	P	(0.0515,0.1079)	(0.0515, 0.0114)
	N	(0.1397, 0.1080)	(0.1397,0.0114)

Table 12. combines the BNE of both interactions, and the standard tag for reviews is deduced. Our objective is to extract a common response from both interactions that provide the sentiment tag for each review. The sentiment tag for R_1 is positive, and R_2 is negative. Similarly, we enable interactions among all reviews in a dataset and tag reviews with sentiments.

Table 9
Rating Type Matrix.

R_1			
R_2		***P***	***N***
	P	(0.2000,0.8000)	(0.2000,0.2000)
	N	(0.8000,0.8000)	(0.8000,0.2000)

Table 6
Combined matrix of Context and Rating.

	R_2				
		PP	*PN*	*NP*	*NN*
R_1	***P***	(0.2463,0.0814)	(0.2463,0.2014)	(1.0231,0.1518)	**(1.0231,0.2718)**
	N	(0.0491,0.0814)	(0.0491,0.2014)	(0.0491,0.1518)	(0.0491,0.2718)

Table 10
Combined matrix of both Context and Rating type matrix.

		R_1			
		PP	*PN*	*NP*	*NN*
R_2	***P***	(0.0103,0.0216)	(0.1703,0.1816)	(0.1703,0.6470)	(0.1703,0.1623)
	N	**(0.6679,0.6676)**	(0.6679,0.1816)	(0.6680,0.6470)	(0.6674,0.1623)

4. Experiments and results

In this section, we discuss the datasets used for experimentations, evaluation metrics, and the performance of the proposed model on the datasets.

4.1. Data collection

We collected three data sets on three domains, viz. food, mobile, and electronic reviews, from the Amazon.com[3] website. Each dataset contains reviews, comments, and their respective ratings. The data statistics of each dataset are listed in Table 13.

4.2. Performance evaluation using metrics

The robustness of the proposed framework was gauged using various evaluation metrics. These include F1-measure, accuracy, precision, etc. It is crucial to analyze a model using a variety of measures because a model may perform well with one evaluation metric but poorly with another. The evaluation metrics used to check the performance of the proposed model are given in Eqs. (9)–(17). Fig. 10. depicts the performance of the proposed model in terms of the different evaluation metrics across the three datasets.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

$$FPR = \frac{FP}{FP + TN} \tag{13}$$

$$FNR = \frac{FN}{TP + FN} \tag{14}$$

$$FDR = \frac{FP}{(TP + FP)} \tag{15}$$

$$F1_{score} = \frac{(2*precision*recall)}{(precision + recall)} \tag{16}$$

$$MCC = \frac{(TP*TN - FP*FN)}{\sqrt{(TP + FP)*(TN + FN)*(FP + TN)*(TP + FN)}} \tag{17}$$

Table 11
Bayesian Nash Equilibrium: Best response of the individual player.

		R_1
		PP
R_2	***N***	**(0.6679,0.6676)**

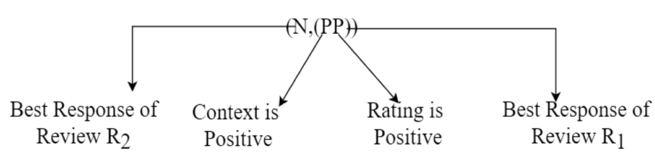

Fig. 9. BNE of interaction 2.

Table 12
Best responses of Interaction 1 and Interaction 2.

	R_1 (4 Star)	R_2 (1 star)
BNE of Interaction 1	*(P)*	*(NN)*
BNE of Interaction 2	*(PP)*	*(N)*
Total Sentiment tag	$P(1+P)$	$N(1+N)$
Deduced sentiment tag for each review	Positive	Negative

Table 13
Data statistics of the three datasets.

Data Set	Language	Positive	Negative
Food reviews	English	629	371
Mobile reviews	English	557	443
Electronic reviews	English	451	549

4.3. Performance of the proposed model with food reviews dataset

We evaluated the performance of the proposed model using the Food reviews dataset. We received an accuracy of 0.923, higher than the current state-of-the-art techniques. To demonstrate the significance of the results, we compared them with the results of models proposed in the literature. Feng and Lin (2016) proposed a sentiment analysis algorithm in which RNN(4 classes) gave the best results. Tran, Trieu, Dao, Nguyen,

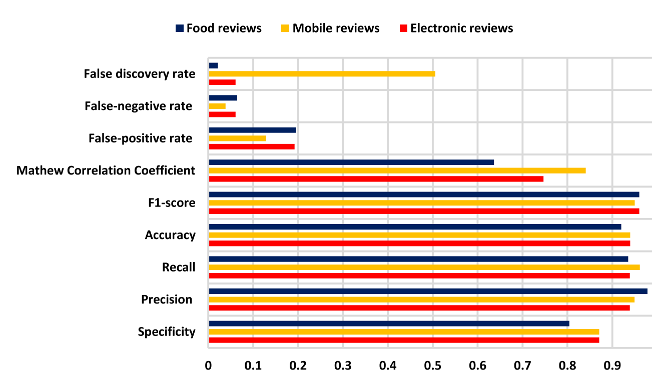

Fig. 10. Performance of the proposed model over the three datasets in terms of different evaluation metrics.

[3] https://jmcauley.ucsd.edu/data/amazon/.

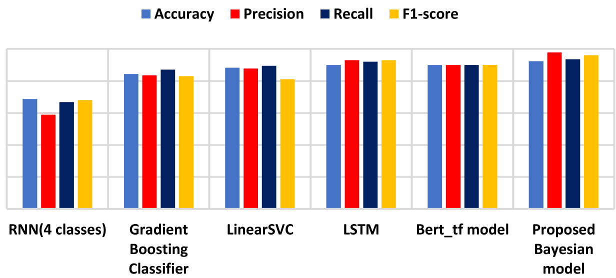

Fig. 11. Comparison of the proposed model's performance with supervised algorithms.

Table 14
Comparison of performance of the proposed model with supervised approaches.

Algorithms	Accuracy	Precision	Recall	F1-score
RNN (Alharbi et al., 2021)	0.758	0.745	0.759	0.749
GRU-based RNN (Alharbi et al., 2021)	0.701	0.632	0.701	0.659
UGRU-based RNN (Alharbi et al., 2021)	**0.906**	**0.906**	**0.907**	**0.903**
UGRNN-based RNN (Alharbi et al., 2021)	0.872	0.872	0.872	0.865
GLSTM-based RNN (Alharbi et al., 2021)	0.883	0.884	0.883	0.877
Proposed Model	**0.940**	**0.953**	**0.962**	**0.950**

and Huynh (2020) also used many supervised models, out of which Gradient Boosting Classifier shows the superlative performance. We also compared the performance with the algorithm proposed by Ahmed et al. (2021). They used supervised approaches in which Linear SVC showed appreciable accuracy. Similarly, Güner, Coyne, and Technology (2019) compared various supervised models for sentiment classification. Out of all, LSTM showed the best performance. Lastly, we compared the performance of the proposed approach with the Bert_tf[4] model. Fig. 11. shows comparisons of the accuracy, precision, recall, and F1-score of different models.

4.4. Performance of the proposed model with movie reviews dataset

We tested the proposed algorithm with the movie reviews dataset. We received state-of-the-art results. To establish the prominence of the results, we compared them with supervised and unsupervised techniques. Alharbi, Alghamdi, Alkhammash, and Al Amri (2021) proposed simple RNN and its four variants for sentiment analysis of the movies dataset. The four variants of RNN were gated recurrent unit (GRNN), update recurrent unit (URNN), Group Long Short-Term Memory Networks (GLSTM), and update gate recurrent unit (UGRNN). The comparison of the results of the proposed model with other supervised approaches is depicted in Table 14.

Table 15
Comparison of performance of the proposed model with unsupervised approaches.

Algorithms	Accuracy	Precision	Recall	F1-score
LDA model (Yiran& Srivastava, 2019)	0.542	0.548	0.566	0.551
Hu and Liu approach (Hu & Liu, 2004)	0.712	0.731	0.696	0.723
Proposed Model	**0.940**	**0.953**	**0.962**	**0.950**

[4] https://github.com/jinudaniel/amazon-fine-food-reviews/blob/master/amazon_reviews_bert_tf.ipynb.

Table 16
Comparison of performance of the Proposed Bayesian Model with the supervised algorithm.

Algorithms	Accuracy	Precision	Recall	F1-score
Maximum entropy Model (Somprasertsri & Lalitrojwong, 2008)	0.743	0.718	0.752	0.735
LSVM classifier (Daniel & Meena, 2021)	0.779	0.782	0.788	0.784
Proposed Bayesian Model	**0.940**	**0.931**	**0.939**	**0.960**

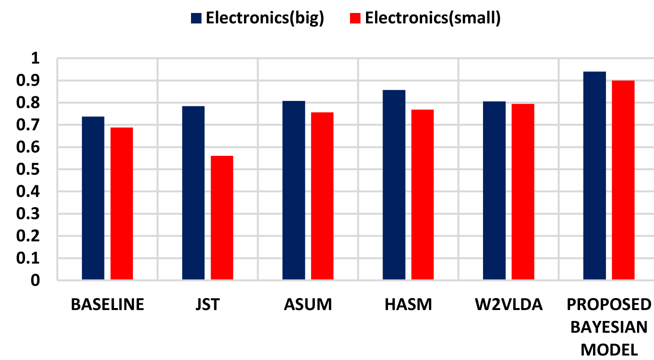

Fig. 12. Comparison of performance of the proposed model with unsupervised approaches.

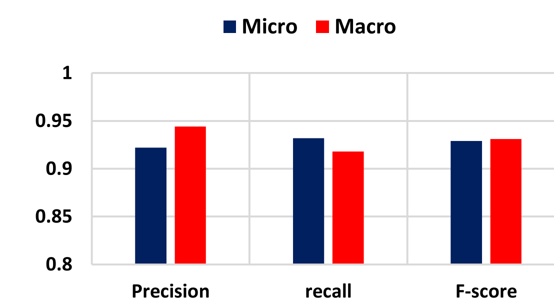

Fig. 13. Macro and Micro averaged performance of the proposed model over three datasets of reviews.

The comparison of the results with the unsupervised approaches using a mobile dataset is tabulated in Table 15. Yiran and Srivastava (2019) clustered topic words with their corresponding probability value for sentiment analysis. The primary objective of Hu and Liu (2004) was to summarize all of a product's customer reviews. To achieve the aim, they performed sentiment analysis of reviews.

4.5. Performance of the proposed model with electronic reviews dataset

The third dataset we used to test the performance of the proposed model was the Electronics reviews dataset. We evaluated the performance and found that the F1-score value is 0.960, higher than the state-of-the-art results. We compared the achieved results with supervised methods proposed by Somprasertsri and Lalitrojwong (2008) and Daniel and Meena (2021). We also compared the results with unsupervised techniques proposed by JST (Lin et al., 2012), ASUM (Jo & Oh, 2011), HASM (Kim et al., 2013), W2VLDA (García-Pablos et al., 2017).

Among the supervised methods, Somprasertsri and Lalitrojwong (2008) used the maximum entropy model to extract product features from online customer reviews and perform the sentiment classification. Daniel and Meena (2021) used the LSVM classifier. The performance scores are mentioned in Table 16. The Proposed Model performs better than both approaches.

The unsupervised Baseline was proposed by Turney and Littman (2003). The approach used polarity seed count, which uses the polarity seed word to assign the most significant proportion to a sentence. W2VLDA used a single seed word per sentiment polarity. JST, ASUM, HASM are the topic modeling-based approaches. We tested the performance of their proposed algorithm on a small dataset with 1000 reviews having 1-star or 5-star ratings and a large dataset of 1000 reviews having 2-star and 4-star ratings. The results of the sentiment categorization comparison are shown in Fig. 12. The proposed model produces comparable results for small datasets and superior results for large datasets. The recorded accuracy for a small dataset is 0.90 and for the big dataset is 0.940.

4.6. Macro and micro evaluation

Macro and micro averages estimate the overall performance of an approach while dealing with different datasets. The overall performance of the proposed model over the three datasets is illustrated in Fig. 13. The formulas for calculation of macro and micro-averaged accuracy, F-score, and recall across the n datasets are given from Eq. (18). to Eq. (23). We received the micro and macro precisions collectively for the three datasets as 0.922 and 0.944, respectively. The values are close to 1.00 symbolizing steady results. Over the three datasets, the macro and micro F1-score values are 0.929 and 0.931, respectively. This verifies that the proposed model produces consistent results when applied to different datasets.

$$Macro-averaged\ Precision = \frac{\sum_{i=1}^{n} Precision_i}{n} \tag{18}$$

$$Macro-averaged\ Recall = \frac{\sum_{i=1}^{n} Recall_i}{n} \tag{19}$$

$$MacroF-score = 2^{*}\frac{Macro\ Presicion^{*}Macro\ Recall}{Macro\ Precision + Macro\ Recall} \tag{20}$$

$$Micro-averaged\ Precision = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + FP_i} \tag{21}$$

$$Micro-averaged\ Recall = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + FN_i} \tag{22}$$

$$MicroF-score = 2^{*}\frac{Micro\ Presicion^{*}Micro\ Recall}{Micro\ Precision + Micro\ Recall} \tag{23}$$

4.7. Validation of proposed Bayesian model through statistical significance

Model validation ensures that the output of a statistical model is sufficiently comparable to the outputs of the data-generation process for the investigation's objectives to be met. Researchers perform validation of their models using statistical significance to validate their results (Ahuja & Sharma, 2021; Joshi et al., 2021; Keshavarz Ghorabaee, Kazimieras Zavadskas, Olfat, & Turskis, 2015, Mladenović, Mitrović, Krstev, & Vitas, 2016). We validated the performance of the proposed model by undertaking two z-tests. One test used two random samples from the same dataset, and the other used two samples from different datasets.

4.7.1. Statistical test within a dataset

We conducted the statistical z-test of two proportions to validate the proposed model's performance. Two samples of different sample sizes were taken from the food reviews dataset. In *sample 1,* we took *497* reviews (N_1). Out of *497* reviews, *461* (X_1) were correctly tagged. Sample proportion of correctly tagged reviews (p_1) = *0.928*. In *sample 2,* we took *252* reviews (N_2). *232* (X_2) were correctly tagged, so the proportion of correctly tagged reviews is *0.921* (p_2). Table 17. summarizes the data.

Table 17
Test statistics of z-test of two proportions of a dataset.

Parameters	Sample 1	Sample 2
Sample Size	497	252
Correctly tagged reviews	461	232
The proportion of correctly tagged reviews	0.928	0.921

Fig. 14. Graphical representation of the critical region of the Hypothesis.

We designed the following null hypothesis (H_0) and alternative hypotheses (H_a) to test the validation of our model.

H_o: proportion of correctly tagged reviews is the same in both samples, $p_1 = p_2$.
H_a: proportion of correctly tagged reviews differs in both samples, $p_1 \neq p_2$.

The value of the pooled proportion is computed using Eq. (24).

$$P = \frac{X_1 + X_2}{N_1 + N_2} = \frac{461 + 232}{497 + 252} = 0.925 \tag{24}$$

A two-tailed test was employed, and a z-test for two population proportions was used. The following Eq. (25) shows the calculation of the z-statistic.

$$z = \frac{p_1 - p_2}{\sqrt{P(1-P)(1/N_1 + 1/N_2)}} = \frac{0.928 - 0.921}{\sqrt{0.925 \times (1 - 0.925)(1/497 + 1/252)}} = 0.344 \tag{25}$$

Fig. 14 shows a graphical representation of the hypothesis's critical region. We failed to reject the null hypothesis H_o. As a result, there is insufficient evidence to assert that the population proportion p_1 differs from p_2 at $\alpha = 0.05$ significance level. This implies that the accuracy of our model is consistent throughout different sample sizes of the same datasets. Thus, the results of the Proposed Bayesian Model for sentiment tagging can be trusted.

4.7.2. Statistical significance within two different datasets

We conducted the z-test for two proportion tests over two different datasets. We took one sample from the Mobile reviews dataset and another from the food reviews dataset. For *sample 1,* we took *100* reviews (N_1). Out of *100* reviews, 94 (X_1) reviews were correctly tagged, so the sample proportion is *0.94* (p_1). Similarly, for *sample 2*, the number of reviews (N_2) = *252.232* (X_2) reviews were correctly tagged, so the sample proportion is *0.921*(p_2). We conducted the Z-test for two population proportions (p_1 and p_2). The data is summarized in Table 18.

We considered the following null hypotheses (H_o) and alternative hypotheses (H_a).

Table 18
Test statistics of Z-test of 2 proportions within the different datasets.

Parameters	Sample 1	Sample 2
Sample Size	100	252
Correctly tagged reviews	94	232
The proportion of Correctly tagged reviews	0.940	0.921

H_0: $p_1 = p_2$, i.e., the proportion of correctly tagged reviews of *sample 1* is equal to correctly tagged of *sample 2*.
H_a: $p_1 \neq p_2$, i.e., the proportion of correctly tagged reviews of *sample 1* is not equal to the correctly tagged of *sample 2*.

The value of the pooled proportion is computed using Eq. (26).

$$P = \frac{X_1 + X_2}{N_1 + N_2} = \frac{94 + 232}{100 + 252} = 0.926 \tag{26}$$

A two-tailed test was employed, and a z-test for two population proportions was used. Eq. (27) was used to calculate the z-statistic.

$$z = \frac{p_1 - p_2}{\sqrt{P(1-P)(1/N_1 + 1/N_2)}} = \frac{0.940 - 0.921}{\sqrt{0.926 \times (1 - 0.926)(1/100 + 1/252)}} = 0.615 \tag{27}$$

Fig. 15 shows the graphical representation of the hypothesis's critical region. Since there is insufficient evidence to assert that the population proportion p_1 differs from p_2 at $\alpha = 0.05$ significance level. Thus, we failed to reject the null hypothesis H_0. This implies that the accuracy of the proposed model is consistent throughout different datasets. This indicates that the results of the proposed model for sentiment tagging are credible over the samples of different datasets.

5. Discussion and error analysis

The proposed model is an initial demonstration of Bayesian games' efficiency for binary classification tasks in NLP. We were successful in demonstrating the utility of Bayesian games for sentiment analysis. This section discusses the proposed work's error rate, computational complexity, and challenges.

1) Evaluation of error rates
Error rate (ERR), also known as misclassification rate, shows the number of incorrectly identified samples from the actual positive and negative classifications sample. It is evaluated using Eq. (28).

$$ERR = \frac{FP + FN}{TP + TN + FN + FP} \tag{28}$$

The error rate performance of several supervised and unsupervised algorithms is depicted in Fig. 16. The LDA model has the highest error rate of 0.458, while LSTM and Bert_tf have nearly equivalent error rates of 0.1. The proposed model has the lowest error rate of 0.077, ensuring that it is more reliable than existing algorithms.

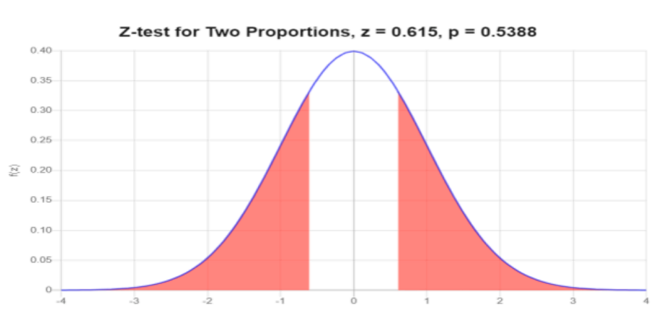

Fig. 15. Graphical representation of the critical region of the above hypothesis.

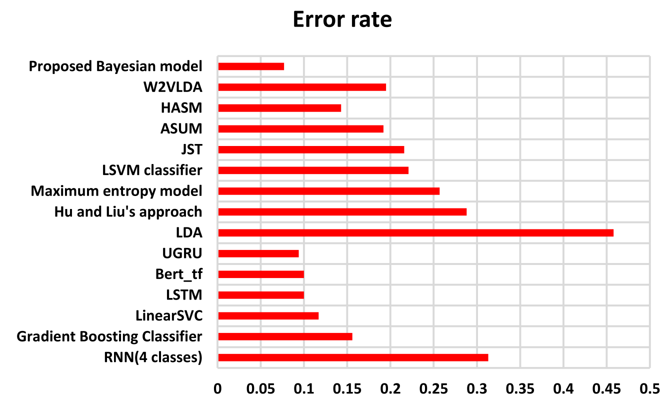

Fig. 16. Comparison of an error rate of the proposed model and the existing approaches.

2) Time and space complexity of the proposed Bayesian model
An algorithm's time and space complexity are essential for assessing its efficiency and effectiveness. This section explains the proposed model's time and space complexity.

Space complexity: The overall space occupied by an algorithm concerning the input size is referred as its space complexity. Let a dataset has n reviews, and each review has W_R POS tagged sentiment-bearing words. The space required to store the dataset's context and rating scores is $2n + 2n = 4n$. The space required to play Bayesian games among all the reviews $= 8n(n\text{-}1)$. So, the total space required is $8n^2 - 8n + 4n + n * W_R$. Thus, the space complexity $= O(n^2)$.

Time Complexity: The amount of time it takes an algorithm to run as a function of the length of the input is referred to as time complexity. It calculates the time required to execute each code statement in an algorithm. Assuming each operation takes almost equal time, we add up the number of operations to find the time complexity of the proposed algorithm. The number of operations to get context score $= n \times W_R$. The number of multiplications between two matrices of order $2 \times 2 = 32$. Number of operations for playing Bayesian games $= 32\times (n(n\text{-}1))$. Thus, the total time complexity $= 32 \times n(n\text{-}1) = O(n^2)$.

3) Performance of the proposed Bayesian model in a general dataset of multiple domains
To check the robustness of the proposed model, we generated two general domain small datasets of two different languages. The first dataset consists of English review comments with diversified domains. The second dataset is the Hindi language dataset of the general domain. The dataset statistics of both datasets are listed in Table 19.

The results after implementing the proposed model on two datasets

Table 19
Statistics of two General domain datasets of different languages.

English Language dataset		Hindi language dataset	
Domains	No. of reviews	Domains	No. of reviews
TripAdvisor reviews	38	Movie reviews	35
Movie reviews	36	Hotel reviews	56
Food reviews	49	Electronics reviews	53
Laptop reviews	29	Zomato and Swiggy reviews	68
Mama earth product reviews	27	Newspaper reviews	59
Beauty product reviews	48		
Mobile reviews	34		
Cloths reviews	25		
Total	**286**	**Total**	**271**

Table 20
Performance of Proposed Bayesian Model over two different language datasets.

Datasets	Accuracy	Precision	Recall	F1-score
General English Dataset	0.866	0.878	0.862	0.859
General Hindi Dataset	0.812	0.812	0.850	0.808

Table 21
Examples where the Proposed Bayesian Model failed to predict accurate results.

S. No.	Reviews
1	*"I do not dislike noodles."*
2	*"The movie is not good for first time watch. "*
3	*"Someone who works as a pizza man does not like pizza?"*
4	*"Pizza is tasty but not at that price."*

are shown in Table 20. The proposed Bayesian Model's performance across multiple domains and languages is encouraging and promising. These additional experiments verified the advantages of the proposed model over other approaches in the literature. The Proposed Bayesian Model can easily adapt to any language. The condition is that we have SentiWordNet or a similar knowledge base to compute the context scores of a review comment in that language. Also, the Proposed Model is domain independent and can be applied to any domain datasets having review comments and ratings for reviews. Also, unlike supervised approaches, the Proposed Bayesian Model does not require training.

4) Challenges in the proposed Bayesian model
The Proposed Bayesian Model is not a full proof model for sentiment classification. We identified the reasons for low performance, which we wish to tackle in future work. During this investigation, the model mislabeled a few of the 3-star reviews. For example: *"The flavors are good. However, I do not see any difference between this and Oaker Oats brand-they are both mushy"*. *"The item over all was fine, but the Banana heads are not just like banana runts they are allot smaller"*. *"I thought it was a great buy, but as noted by Scott D., the molasses content was not up to par"*. The reviews in the instances above have a three-star rating, and the proposed algorithm failed to tag them correctly. A possible reason for the failure is that the review comments included an equal quantity of positive and negative sentiments. Furthermore, because the star rating was 3, the rating failed to serve any purpose. We wish to address the neutrality or ambivalence of sentiment classification using Bayesian games as the next step in future research.

Next challenge is that the proposed methodology is unable to categorize the constructs like negation phrases. Negative words influence the sentiment orientation of other words in a phrase. Negation terms include not, no, never, cannot, shouldn't, wouldn't, and so on. When a negation appears in a sentence, it is critical to determine the words impacted by the negation. Few of the examples with negation are listed in Table 21. However, these roadblocks can be tackled, and we intend to handle negation in our future work.

6. Conclusion and future work

The proposed work demonstrates mathematical optimization techniques, such as Game Theory, to perform sentiment analysis of reviews. The proposed model is a novel unsupervised technique for sentiment tagging of reviews. We evaluated the effectiveness of our model on three review datasets and compared the results with the state-of-the-art approaches. The proposed model outperformed the state-of-the-art supervised and unsupervised techniques. We also statistically cross-validated the performance of the proposed methodology. The presented model paved the path for Bayesian games to be applied to various NLP tasks such as WSD and query expansion, among others. One of the advantages of the proposed model is that it can be easily adapted for sentiment classification of low-resource languages like Hindi, Bengali, Urdu, etc., with minimal variations. We understand that there is an enormous scope for improvement for us. We intend to extend the proposed model for multi-class classification of sentiments as the next step. The challenges like negation handling is also alluring and we intend to take up this task in future as well.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have attached the code file pdf in the attached file step and link for the datasets are given in the manuscript

References:

Abapour, S., Mohammadi-Ivatloo, B., & Tarafdar Hagh, M. (2020). A Bayesian game theoretic based bidding strategy for demand response aggregators in electricity markets. *Sustainable Cities and Society, 54*, Article 101787. https://doi.org/10.1016/J.SCS.2019.101787

Agarwal, B., Mittal, N., Bansal, P., & Garg, S. (2015). Sentiment analysis using common-sense and context information. *Computational Intelligence and Neuroscience, 2015*. https://doi.org/10.1155/2015/715730

Ahmad, A., & Ahmad, T. (2019). A Game Theory Approach for Multi-document Summarization. *Arabian Journal for Science and Engineering, 44*(4), 3655–3667. https://doi.org/10.1007/S13369-018-3619-Y

Ahmed, H. M., Awan, M. J., Khan, N. S., Yasin, A., Muhammad, H., & Shehzad, F. (2021). Sentiment analysis of online food reviews using big data analytics. *Papers. Ssrn. Com, 20*(2), 827–836. https://doi.org/10.17051/ilkonline.2021.02.93

Ahuja, R., & Sharma, S. C. (2021). Transformer-Based Word Embedding With CNN Model to Detect Sarcasm and Irony. *Arabian Journal for Science and Engineering, 1–14*. https://doi.org/10.1007/S13369-021-06193-3/TABLES/8

Akhtar, M. S., Ekbal, A., & Cambria, E. (2020). How Intense Are You? Predicting Intensities of Emotions and Sentiments using Stacked Ensemble [Application Notes]. *IEEE Computational Intelligence Magazine, 15*(1), 64–75. https://doi.org/10.1109/MCI.2019.2954667

Aldoğan, D., & Yaslan, Y. (2017). A comparison study on active learning integrated ensemble approaches in sentiment analysis. *Computers & Electrical Engineering, 57*, 311–323. https://doi.org/10.1016/J.COMPELECENG.2016.11.015

Alharbi, N. M., Alghamdi, N. S., Alkhammash, E. H., & Al Amri, J. F. (2021). Evaluation of Sentiment Analysis via Word Embedding and RNN Variants for Amazon Online Reviews. *Mathematical Problems in Engineering, 2021*. https://doi.org/10.1155/2021/5536560

Anatomy, A. N., & Mahadevan, B. (2000). Business Models for Internet-Based E-Commerce. *California CMR, 42*.

Arsenyan, J., Büyüközkan, G., & Feyzioǧlu, O. (2015). Modeling collaboration formation with a game theory approach. *Expert Systems with Applications, 42*(4), 2073–2085. https://doi.org/10.1016/j.eswa.2014.10.010

Atoum, I. (2020). A novel framework for measuring software quality-in-use based on semantic similarity and sentiment analysis of software reviews. *Journal of King Saud University - Computer and Information Sciences, 32*(1), 113–125. https://doi.org/10.1016/j.jksuci.2018.04.012

Barfar, A. (2022). A linguistic/game-theoretic approach to detection/explanation of propaganda. *Expert Systems with Applications, 189*, Article 116069. https://doi.org/10.1016/J.ESWA.2021.116069

Basiri, M. E., Nemati, S., Abdar, M., Cambria, E., & Acharya, U. R. (2021). ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis. *Future Generation Computer Systems, 115*, 279–294. https://doi.org/10.1016/J.FUTURE.2020.08.005

Berka, P. (2020). Sentiment analysis using rule-based and case-based reasoning. *Journal of Intelligent Information Systems, 55*(1), 51–66. https://doi.org/10.1007/S10844-019-00591-8/TABLES/1

Biaou, B. O. S., Oluwatope, A. O., Odukoya, H. O., Babalola, A., Ojo, O. E., & Sossou, E. H. (2020). Ayo game approach to mitigate free riding in peer-to-peer networks. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2020.09.015

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems, 226*. https://doi.org/10.1016/J.KNOSYS.2021.107134

Bu, Z., Li, H., Cao, J., Wu, Z., & Zhang, L. (2016). Game theory based emotional evolution analysis for Chinese online reviews. *Knowledge-Based Systems, 103*, 60–72. https://doi.org/10.1016/j.knosys.2016.03.026

Cao, J., Xu, X., Yin, X., & Pan, B. (2022). A risky large group emergency decision-making method based on topic sentiment analysis. *Expert Systems with Applications, 195*, Article 116527. https://doi.org/10.1016/J.ESWA.2022.116527

Chiha, R., Ayed, M. B., & Pereira, C. da C. (2022). A complete framework for aspect-level and sentence-level sentiment analysis. *Applied Intelligence, 1–19*. https://doi.org/10.1007/S10489-022-03279-9/TABLES/11

Contratres, F. G., Alves-Souza, S. N., Filgueiras, L. V. L., & DeSouza, L. S. (2018). Sentiment analysis of social network data for cold-start relief in recommender systems. *Advances in Intelligent Systems and Computing, 746*, 122–132. https://doi.org/10.1007/978-3-319-77712-2_12

Da'U, A., & Salim, N. (2019). Sentiment-Aware Deep Recommender System with Neural Attention Networks. *IEEE Access, 7*, 45472–45484. https://doi.org/10.1109/ACCESS.2019.2907729

Dahiya, A., & Gupta, B. B. (2021). A reputation score policy and Bayesian game theory based incentivized mechanism for DDoS attacks mitigation and cyber defense. *Future Generation Computer Systems, 117*, 193–204. https://doi.org/10.1016/J.FUTURE.2020.11.027

Daniel, D. A. J., & Meena, M. J. (2021). A Novel Sentiment Analysis for Amazon Data with TSA based Feature Selection. *Scalable Computing: Practice and Experience, 22*(1), 53–66. https://doi.org/10.12694/scpe.v22i1.1839

Donadi, M. (2018). *A System for Sentiment Analysis of Online-Media with TensorFlow*. 1–44. https://reposit.haw-hamburg.de/bitstream/20.500.12738/8446/1/Bachelorarbeit_Donadi_final.pdf.

Dwivedi, A., & Pant, R. P. (2019). An algorithmic implementation of entropic ternary reduct soft sentiment set (ETRSSS) using soft computing technique on big data sentiment analysis (BDSA) for optimal selection of a decision based on real-time update in online reviews. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2019.10.001

Esuli, A., & Sebastiani, F. (2006). SENTIWORDNET: A publicly available lexical resource for opinion mining. *Proceedings of the 5th International Conference on Language Resources and Evaluation*.

Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data, 2*(1), 1–14. https://doi.org/10.1186/S40537-015-0015-2/FIGURES/9

Feng, H., & Lin, R. (2016). Sentiment Classification of Food Reviews. *Arxiv.Org*. http://arxiv.org/abs/1609.01933.

Fiok, K., Karwowski, W., Gutierrez, E., & Wilamowski, M. (2021). Analysis of sentiment in tweets addressed to a single domain-specific Twitter account: Comparison of model performance and explainability of predictions. In *Expert Systems with Applications* (Vol. 186). doi:10.1016/j.eswa.2021.115771.

García-Pablos, A., Cuadros, M., & Rigau, G. (2017). W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis. *Expert Systems with Applications, 91*, 127–137. https://doi.org/10.1016/j.eswa.2017.08.049

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications, 40*(16), 6266–6282. https://doi.org/10.1016/j.eswa.2013.05.057

Güner, L., Coyne, E., Technology, J. S.-B. D. in M., & 2019, U. (2019). *Sentiment analysis for Amazon.com reviews*.

Hossain, N., Bhuiyan, M. R., Tumpa, Z. N., & Hossain, S. A. (2020). Sentiment Analysis of Restaurant Reviews using Combined CNN-LSTM. In *2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020*. doi:10.1109/ICCCNT49239.2020.9225328.

Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. In *Proceedings of the National Conference on Artificial Intelligence*, 755–760.

Huang, D. W., Yang, L. X., Li, P., Yang, X., & Tang, Y. Y. (2021). Developing Cost-Effective Rumor-Refuting Strategy through Game-Theoretic Approach. *IEEE Systems Journal, 15*(4), 5034–5045. https://doi.org/10.1109/JSYST.2020.3020078

Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences, 30*(4), 330–338. https://doi.org/10.1016/j.jksues.2016.04.002

Ileri, M., & Turan, M. (2021). Sentiment Analysis of Meeting Room. In *HORA 2021 - 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications, Proceedings*. doi:10.1109/HORA52670.2021.9461354.

Jain, M., Gayathri, B., Ranjan, R., & Rohit. (2022). SentiGames- A Game Theoretic Approach To Sentiment Analysis. In *7th International Conference on Communication and Electronics Systems, ICCES 2022 - Proceedings*, 963–968. doi:10.1109/ICCES54183.2022.9835933.

Jain, L., Katarya, R., & Sachdeva, S. (2020). Recognition of opinion leaders coalitions in online social network using game theory. *Knowledge-Based Systems, 203*, Article 106158. https://doi.org/10.1016/j.knosys.2020.106158

Jain, G., & Lobiyal, D. K. (2022). Word Sense Disambiguation using Cooperative Game Theory and Fuzzy Hindi WordNet based on ConceptNet. *Transactions on Asian and Low-Resource Language Information Processing, 21*(4), 1–25. https://doi.org/10.1145/3502739

Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review, 41*, Article 100413. https://doi.org/10.1016/J.COSREV.2021.100413

Jain, M., Suvarna, A., & Jain, A. (2021). An evolutionary game theory based approach for query expansion. *Multimedia Tools and Applications, 2021*, 1–25. https://doi.org/10.1007/S11042-021-11297-X

Jeon, J. H., Xia, R., & Liu, Y. (2011). Sentence level emotion recognition based on decisions from subsentence segments. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 4940–4943*. https://doi.org/10.1109/ICASSP.2011.5947464

Jindal, K., & Aron, R. (2021). A systematic study of sentiment analysis for social media data. *Materials Today: Proceedings*. https://doi.org/10.1016/J.MATPR.2021.01.048

Jo, Y., & Oh, A. (2011). Aspect and sentiment unification model for online review analysis. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining, WSDM 2011* (pp. 815–824). https://doi.org/10.1145/1935826.1935932

Karthik, R. V., & Ganapathy, S. (2021). A fuzzy recommendation system for predicting the customers interests using sentiment analysis and ontology in e-commerce. *Applied Soft Computing, 108*. https://doi.org/10.1016/j.asoc.2021.107396

Keshavarz Ghorabaee, M., Kazimieras Zavadskas, E., Olfat, L., & Turskis, Z. (2015). Multi-criteria inventory classification using a new method of evaluation based on distance from average solution (EDAS). *Content. Iospress. Com, 26*(3), 435–451.

Kim, S., Zhang, J., Chen, Z., Oh, A., & Liu, S. (2013). A hierarchical aspect-sentiment model for online reviews. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence, AAAI 2013* (pp. 526–533). https://doi.org/10.1609/aaai.v27i1.8700

Li, M., Chen, L., Zhao, J., & Li, Q. (2021). Sentiment analysis of Chinese stock reviews based on BERT model. *Applied Intelligence, 51*(7), 5016–5024. https://doi.org/10.1007/S10489-020-02101-8/TABLES/5

Li, L., Johnson, J., Aarhus, W., & Shah, D. (2022). Key factors in MOOC pedagogy based on NLP sentiment analysis of learner reviews: What makes a hit. *Computers & Education, 176*, Article 104354. https://doi.org/10.1016/J.COMPEDU.2021.104354

Liang, B., Su, H., Gui, L., Cambria, E., & Xu, R. (2022). Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowledge-Based Systems, 235*, Article 107643. https://doi.org/10.1016/J.KNOSYS.2021.107643

Lin, C., He, Y., Everson, R., & Rüger, S. (2012). Weakly supervised joint sentiment-topic detection from text. *IEEE Transactions on Knowledge and Data Engineering, 24*(6), 1134–1145. https://doi.org/10.1109/TKDE.2011.48

Liu, F., & Li, M. (2019). A game theory-based network rumor spreading model: Based on game experiments. *International Journal of Machine Learning and Cybernetics, 10*(6), 1449–1457. https://doi.org/10.1007/S13042-018-0826-5/FIGURES/5

Liu, N., & Shen, B. (2020). ReMemNN: A novel memory neural network for powerful interaction in aspect-based sentiment analysis. *Neurocomputing, 395*, 66–77. https://doi.org/10.1016/j.neucom.2020.02.018

Mee, A., Homapour, E., Chiclana, F., & Engel, O. (2021). Sentiment analysis using TF–IDF weighting of UK MPs' tweets on Brexit. *Knowledge-Based Systems, 228*, Article 107238. https://doi.org/10.1016/J.KNOSYS.2021.107238

Mladenović, M., Mitrović, J., Krstev, C., & Vitas, D. (2016). Hybrid sentiment analysis framework for a morphologically rich language. *Journal of Intelligent Information Systems, 46*(3), 599–620. https://doi.org/10.1007/s10844-015-0372-5

Muthumanickam, K., & Ilavarasan, E. (2015). Optimization of rootkit revealing system resources - A game theoretic approach. *Journal of King Saud University - Computer and Information Sciences, 27*(4), 386–392. https://doi.org/10.1016/j.jksuci.2014.10.004

Narahari, Y., Narayanam, R., Garg, D., & Prakash, H. (2009). Foundations of Mechanism Design. *Game Theoretic Problems in Network Economics and Mechanism Design Solutions, 1–131*. https://doi.org/10.1007/978-1-84800-938-7_2

Nguyen, D. N., Phan, T. T., & Do, P. (2021). Embedding knowledge on ontology into the corpus by topic to improve the performance of deep learning methods in sentiment analysis. *Scientific Reports, 11*(1). https://doi.org/10.1038/s41598-021-03011-6

Ozyurt, B., & Akcayol, M. A. (2021). A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. *Expert Systems with Applications, 168*. https://doi.org/10.1016/j.eswa.2020.114231

Pandesenda, A. I., Yana, R. R., Sukma, E. A., Yahya, A., Widharto, P., & Hidayanto, A. N. (2020). Sentiment Analysis of Service Quality of Online Healthcare Platform Using Fast Large-Margin. In *Proceedings - 2nd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2020* (pp. 121–125). https://doi.org/10.1109/ICIMCIS51567.2020.9354295

Perikos, I., Kardakis, S., & Hatzilygeroudis, I. (2021). Sentiment analysis using novel and interpretable architectures of Hidden Markov Models[Formula presented]. *Knowledge-Based Systems, 229*, Article 107332. https://doi.org/10.1016/j.knosys.2021.107332

Rani, R., & Lobiyal, D. K. (2020). Performance evaluation of text-mining models with Hindi stopwords lists. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2020.03.003

Ray, P. P. (2018). A survey on Internet of Things architectures. *Journal of King Saud University - Computer and Information Sciences, 30*(3), 291–319. https://doi.org/10.1016/j.jksuci.2016.10.003

Ruseti, S., Sirbu, M. D., Calin, M. A., Dascalu, M., Trausan-Matu, S., & Militaru, G. (2020). Comprehensive exploration of game reviews extraction and opinion mining using nlp techniques. *Advances in Intelligent Systems and Computing, 1041*, 323–331. https://doi.org/10.1007/978-981-15-0637-6_27/COVER

Sarkar, D. (2019). Text Analytics with Python. *Text Analytics with Python*. https://doi.org/10.1007/978-1-4842-4354-1

Saxena, A., Mangal, M., & Jain, G. (2021). *KeyGames: A Game Theoretic Approach to Automatic Keyphrase Extraction*. 2037–2048. doi:10.18653/v1/2020.coling-main.184.

Sharma, R., & Spaan, M. T. J. (2012). Bayesian-Game-Based Fuzzy Reinforcement Learning Control for Decentralized POMDPs. *IEEE Transactions on Computational Intelligence and AI in Games, 4*(4). https://doi.org/10.1109/TCIAIG.2012.2212279

Somprasertsri, G., & Lalitrojwong, P. (2008). A maximum entropy model for product feature extraction in online customer reviews. In *2008 IEEE International Conference on Cybernetics and Intelligent Systems, CIS 2008*, 575–580. doi:10.1109/ICCIS.2008.4670882.

Song, M., Park, H., Shin, & Shik, K. (2019). Attention-based long short-term memory network using sentiment lexicon embedding for aspect-level sentiment analysis in Korean. *Information Processing & Management, 56*(3), 637–653. https://doi.org/10.1016/J.IPM.2018.12.005

Swathi, T., Kasiviswanath, N., & Rao, A. A. (2022). An optimal deep learning-based LSTM for stock price prediction using twitter sentiment analysis. *Applied Intelligence, 1–14*. https://doi.org/10.1007/S10489-022-03175-2/FIGURES/10

Tran, P. Q., Trieu, N. T., Dao, N. V., Nguyen, H. T., & Huynh, H. X. (2020). Effective opinion words extraction for food reviews classification. *International Journal of Advanced Computer Science and Applications, 11*(7), 421–426. https://doi.org/10.14569/IJACSA.2020.0110755

Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems, 21* (4), 315–346. https://doi.org/10.1145/944012.944013

Usama, M., Ahmad, B., Song, E., Hossain, M. S., Alrashoud, M., & Muhammad, G. (2020). Attention-based sentiment analysis using convolutional and recurrent neural network. *Future Generation Computer Systems, 113*, 571–578. https://doi.org/10.1016/J.FUTURE.2020.07.022

Vashishtha, S., & Susan, S. (2019a). Fuzzy rule based unsupervised sentiment analysis from social media posts. In *Expert Systems with Applications* (Vol. 138). doi:10.1016/j.eswa.2019.112834.

Vashishtha, S., & Susan, S. (2021). Highlighting keyphrases using senti-scoring and fuzzy entropy for unsupervised sentiment analysis. In *Expert Systems with Applications* (Vol. 169). doi:10.1016/j.eswa.2020.114323.

Vashishtha, S., & Susan, S. (2019). Fuzzy rule based unsupervised sentiment analysis from social media posts. *Expert Systems With Applications, 138*, Article 112834. https://doi.org/10.1016/j.eswa.2019.112834

Vincent, T. L., & Brown, J. S. (2005). Evolutionary game theory, natural selection, and Darwinian dynamics. *Evolutionary Game Theory, Natural Selection, and Darwinian Dynamics*. https://doi.org/10.1017/CBO9780511542633

von Neumann, J., & Morgenstern, O. (2007). Theory of games and economic behavior (60th anniversary commemorative edition). In *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*, *9781400829460*, 1–741. doi:10.1515/9781400829460/MACHINEREADABLECITATION/BIBTEX.

Wu, J. L., & Chung, W. Y. (2022). Sentiment-based masked language modeling for improving sentence-level valence–arousal prediction. *Applied Intelligence, 1–17*. https://doi.org/10.1007/S10489-022-03384-9/TABLES/7

Wu, J., Ma, X., Chiclana, F., Liu, Y., & Wu, Y. (2022). A consensus group decision making method for hotel selection with online reviews by sentiment analysis. *Applied Intelligence, 52*(9), 10716–10740. https://doi.org/10.1007/S10489-021-02991-2/FIGURES/7

Wu, H., Zhang, Z., Shi, S., Wu, Q., & Song, H. (2022). Phrase dependency relational graph attention network for Aspect-based Sentiment Analysis. *Knowledge-Based Systems, 236*, Article 107736. https://doi.org/10.1016/J.KNOSYS.2021.107736

Xia, S., Lin, F., Chen, Z., Tang, C., Ma, Y., & Yu, X. (2020). A Bayesian Game Based Vehicle-to-Vehicle Electricity Trading Scheme for Blockchain-Enabled Internet of Vehicles. *IEEE Transactions on Vehicular Technology, 69*(7), 6856–6868. https://doi.org/10.1109/TVT.2020.2990443

Xiao, Y., Li, W., Qiang, S., Li, Q., Xiao, H., & Liu, Y. (2022). A Rumor & Anti-Rumor Propagation Model Based on Data Enhancement and Evolutionary Game. *IEEE Transactions on Emerging Topics in Computing, 10*(2), 690–703. https://doi.org/10.1109/TETC.2020.3034188

Xie, H., Lin, W., Lin, S., Wang, J., & Yu, L. C. (2021). A multi-dimensional relation model for dimensional sentiment analysis. *Information Sciences, 579*, 832–844. https://doi.org/10.1016/J.INS.2021.08.052

Yang, Y., Ma, Y., Zhang, J., Gao, X., & Xu, M. (2020). Attpnet: Attention-based deep neural network for 3D point set analysis. *Sensors (Switzerland), 20*(19), 1–20. https://doi.org/10.3390/s20195455

Yiran, Y., & Srivastava, S. (2019). Aspect-based Sentiment Analysis on mobile phone reviews with LDA. *ACM International Conference Proceeding Series, 101–105*. https://doi.org/10.1145/3340997.3341012

Zhang, Z., & Singh, M. P. (2014). ReNew: A semi-supervised framework for generating domain-specific lexicons and sentiment analysis. In *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, *1*, 542–551. doi:10.3115/v1/p14-1051.

Zhao, A., & Yu, Y. (2021). Knowledge-enabled BERT for aspect-based sentiment analysis. *Knowledge-Based Systems, 227*, Article 107220. https://doi.org/10.1016/J.KNOSYS.2021.107220