Received <day> <Month>, <year>; | Revised <day> <Month>, <year>; | Accepted <day> <Month>, <year>

DOI: xxx/xxxx

ARTICLE TYPE

Valuating Requirements Arguments in the online user's forum for requirements decision-making: The CrowdRE-VArg Framework

Javed Ali Khan*[1] | Affan Yasin[2] | Rubia Fatima[2] | Danish Vasan[2] | Arif Ali Khan[3] | Abdul Wahid[4]

[1]Department of software engineering, University of Science and Technology Bannu, at Bannu, Pakistan
[2]School of software, Tsinghua University, Beijing, P.R.China
[3]M3S Empirical Software Engineering Research Unit,University of Oulu, Oulu, Finland
[4]Department of Computer Science, University of Sciences and Technology Bannu, at Bannu, Pakistan

Correspondence
*Javed Ali Khan. Email: engr_javed501@yahoo.com

Abstract

User forums enable a large population of crowd-users to publicly share their experience, useful thoughts, and concerns about the software applications in the form of user reviews. Recent research studies have revealed that end-user reviews contain rich and pivotal sources of information for the software vendors and developers that can help undertake software evolution and maintenance tasks. However, such user-generated information is often fragmented, with multiple viewpoints from various stakeholders involved in the ongoing discussions in the Reddit forum. In this paper, we proposed a Crowd-based Requirements Engineering by Valuation Argumentation (CrowdRE-VArg) approach that analyses the end-users discussion in the Reddit forum and identifies conflict-free new features, design alternatives, or issues, and reach a rationale-based requirements decision by gradually valuating the relative strength of their supporting and attacking arguments. The proposed approach helps to negotiate the conflict over the new features or issues between the different crowd-users on the run by finding a settlement that satisfies the involved crowd-users in the ongoing discussion in the Reddit forum using argumentation theory. For this purpose, we adopted the bipolar gradual valuation argumentation framework, extended from the abstract argumentation framework and abstract valuation framework. The automated CrowdRE-VArg approach is illustrated through a sample crowd-users conversation topic adopted from the Reddit forum about Google Map mobile application. Finally, we applied natural language processing and different machine learning algorithms to support the automated execution of the CrowdRE-VArg approach. The results demonstrate that the proposed CrowdRE-VArg approach works as a proof-of-concept and automatically identifies prioritized requirements-related information for software engineers.

KEYWORDS:
Argumentation, requirements, Reddit forum, natural language processing, machine learning, new features

1 | INTRODUCTION

Recently, social media such as Reddit forum, Twitter, and app stores have grown exponentially in use and its impact on the market-driven software products, aiming at different stakeholders distributed geographically to enhance the current software features and the overall quality. This research study analyses end-user reviews in the Reddit forum and highlights their importance for software evolution. Also, reviews from the relevant social media platforms and the Reddit forum serve as a communication channel between the software developers and the crowd-users where end-users can provide useful information to help software app developers accomplish certain maintenance tasks [1,2]. For example, accommodating new features [3,4], capturing cross-domain requirements [5], bug identification [6], domain knowledge, or improving the non-function aspects [7]. However, because of the large number of reviews posted daily to the Reddit forum, the contributors of the market-driven software's face a major challenge in incorporating the user reviews into the system under development [8]. Also, end-user comments in the Reddit forum possess certain conflicts because of their hierarchical structure in the forum [3], where one comment comes as a reply to their immediate parent comment. Thus highlighting the importance of rationale management for the market-driven software development.

Rationale management foresees how to capture and record the justification for why a certain decision was taken [9]. A series of software decisions are taken when designing and developing a software application [10]. These decisions, alternatives considered, and argumentation recorded for the agreed choices (or rejected) result in software rationale [11]. However, this rationale information is not explicitly recorded during the requirements and development phase. Rather, rationale information remains hidden or implicit in the software developer's mind or intertwined with software development artefacts [12]. Recently, in software engineering literature, requirements and design rationale management remain the active research area [13]. In the requirements engineering (RE) phase, requirements engineers or software developers of market-based software applications make various decisions to finalize the software features or requirements [10]. Therefore, performing rationale management is a natural expectation to capture justification for the decision taken during the RE phase [14]. Although there seems to be widespread agreement about the significance of software rationale, it is still believed to be rarely captured in practice [15] during requirements and design phases [16]. It is hard to capture software rationale due to its high cost, and the agile methodologies in practice, where documentation is kept to a minimum, cause the software developer's tacit knowledge to remain unexplored [17,11]. An alternative solution is to explore if the rationale information is mined from the existing documentation, web forums, and the existing software artefacts [11]. The relevant social media platforms and user forums (Reddit) provide another channel to interact with diverse users and collect rich rationale information [18,9]. Researchers have recently set out to mine intuitive requirements and design rationale information from the user forums [9,3], issue tracking system [10,11], Twitter [19], and user reviews [20].

Furthermore, researchers have recently utilized argumentation theory in the RE domain, which is widely accepted and supported by the RE research community. Whereas argumentation theory is that an argument is trusted if it can defend itself successfully against the attacking arguments [21]. Such as, it is utilized to eradicate requirements ambiguities during interviews [22], recognize conflict-free user requirements by removing inconsistencies [23], identify tacit knowledge during requirements elicitation [24], and bring most important issues upfront to the development team [25,26]. Besides, some researchers have started using argumentation theory and machine learning (ML) algorithms to mine user comments from the Reddit forum [27] and issue tracking systems (ITSs) [8] to identify conflict-free new features, issues and usability requirements. It inspires us to conduct an exploratory study on a user forum like (Reddit.com) to identify the strength values of requirements-related arguments that would help requirements engineers in decision-making to finalize conflict-free features and issues. By manually analyzing the discussion topics in the Reddit forum, it is identified that most of the user comments possess an argumentative structure and contain rich requirements-related information [28]. Also, during the ongoing interaction in the Reddit forum, the crowd-users could suggest a new feature, issue, or alternative solution in reply to the fellow user comment or the main discussion topic. Considering the user discussion flow and their nested comment-reply structure, we hypothesize that we could identify the strength of user comments (new feature, issue, or the main discussion topic) by valuating the strength of their corresponding supporting or attacking arguments using aggregation and combination functions [29], natural language processing (NLP) tools, and ML classifiers.

Motivated and inspired by the discussions mentioned above, this research study mainly reports on a three-fold study aiming to leverage the Reddit forums as a pivotal source of useful crowd-user feedback information for software maintenance and evolution. Particularly, we employ the bipolar gradual valuation argumentation (GVBA) framework [30], which is extended from the abstract argumentation framework (AAF) [21], and abstract valuation framework [31] to identify conflict-free new features or issues and reach a rationale requirements decision by gradually valuating the relative strength of arguments (their supporting and attacking arguments) in the Reddit forum. For this purpose, we proposed the CrowdRE-VArg approach that helps negotiate the conflicts over the new features or issues between the different crowd-users on the run by finding a settlement that satisfies the involved crowd-users in the ongoing discussion on the Reddit forum. The proposed CrowdRE-VArg framework is inspired from the Gradual Valuation For Abstract Argumentation Framework (GVAAF) [31], GVBA [30], Quantitative argumentation debate framework (QuAD) [32], Extended social abstract argumentation framework (ESAAF) [33], and discontinuity-free Quantitative argumentation debate framework (DF-QuAD) [29]. Secondly, to automate the CrowdRE-VArg approach, we developed an algorithm that automatically identifies conflict-free new features and issues by gradually valuating their supporting and attacking arguments to reach a rationale-based requirements decision. Finally, to scale the proposed CrowdRE-VArg approach, we employ different ML algorithms that classify user comments into new features, issues, and claim rationale elements. Then employee sentiments analysis approach over the user comments identified as claim rationale element to identify end-user opinions (supporting, attacking, or neutral) about the new features, issues, or the main discussion topics. Furthermore, the proposed CrowdRE-VArg approach is elaborated with a crowd-users discussion topic on the google maps mobile application from the Reddit forum.

The main structure of the paper is as follows: section 2 discusses the research papers identified using literature survey, key related work, and comparison with existing literature. Section 3 introduces the study background, research questions and proposed methodological details. Section 4 discusses the proposed CrowdRE-VArg approach to negotiate conflicts between the crowd-users and achieve requirements decision-making. In section 5, the classification of rationale and requirements-related elements are elaborated, and experimental results are explained. Section 6 elaborates on automated argumentation-based requirements decision making with the help of a case study. In section 7, we discuss the research findings. Similarly, section 8 explains the threats to validity, and section 9 concludes the research paper and highlights the future directions.

2 | RELATED WORK

This section highlighted the related work on RE with argumentation and rationale mining. In software engineering literature, the most common and prominent argumentation techniques used to identify conflicting viewpoints are AAF[21] and[34].

2.1 | Literature Survey

The Crowd literature survey identifies relevant research papers on the topic under discussion. For this purpose, we used snowballing sampling approach[35] that identifies research papers found in the reference section of the related research manuscripts, referred to as backward snowballing. Also, identifies and includes research papers cited in the related research articles, called forward snowballing. A similar research approach has been adopted by Khan et al.[36] along with experimental work in the domain of software engineering. Steadily, the sample size of the shortlisted research studies increases as more citations and references are explored on argumentation-based CrowdRE. The relevant literature studies are listed to extract the related work on argumentation and rationale-based CrowdRE that would help recover the differences and advantages of the proposed CrowdRE-VArg approach. We emphasise the literature studies that discuss and elaborate on argumentation and rationale-based RE and software approaches. The search string used to identify the related literature work is Argumentation-based AND (Requirements Engineering OR RE OR CrowdRE) AND (software OR requirements) rationale management. The Google Scholar search engine is used to search the literature studies. It provides an interface to search the scholarly research articles in various popular digital libraries, i.e., IEEE Xplore, Wiley online library, Springer, ACM, etc. Also, main software and RE-related journals, conferences, and workshops that meet the developed search criteria are explored to ensure the inclusion of important papers, e.g., software: practice and experience, RE conference, journal of software: process and evolution, RE Journal, CAiSE, IET software, CrowdRE workshop, IST, and JSS. It gives confidence that no relevant journals, conferences, workshops, and any important digital library have been missed. Mainly, the first two authors of the manuscript were involved in the literature data collection. Moreover, conflicts between authors on selecting literature papers at any point have been resolved by discussing them with the remaining manuscript authors. The inclusion criteria for identifying related works are: the manuscript must be written in English and published in peer-reviewed journals, conferences, a refereed thesis, or a book chapter.

We identified 86 research papers using forward and backwards snowballing by this criteria. Then using a manual selection process, we excluded research papers written in languages other than English or published in unrecognized venues. To accomplish it, we read the manuscript title, its abstract, or the overall article if its relevancy is still uncertain. In total, 78 papers were selected, amongst which 23 research papers are directly related to the proposed CrowdRE-VArg approach that is discussed and compared in the related section. In contrast, 26 articles selected discusses CrowdRE concepts in general and are indirectly related to the research topic. While 20 papers discussed argumentation theory and mining that inspired the proposed CrowdRE-VArg approach, seven research papers are supporting articles that discuss useful links and other concepts used in the proposed method. Additionally, the third and fourth authors validate the final list of the research papers for the literature review to ensure no relevant research paper has been missed. Based on the SLR below, we discussed the related work using argumentation theory and rationale in requirements and software engineering. Later, we compare the proposed approach with the existing approaches and elaborate in Table 1 on the improvements and effectiveness achieved by the proposed CrowdRE-VArg approach compared to the existing approaches.

2.2 | Requirements Engineering with Argumentation theory

Haley et al.[37] first introduce argumentation in security requirements by evaluating system security requirements. While Franqueira et al.[38] proposed a risk assignment methodology, RISA, by extending Haley's work to identify rebuttals and mitigation for security requirements. Kovacs et al.[39] extend Franqueira's work by validating the RISA framework with a complex BitMessage chat application to analyze security requirements. Ionita et al.[40] proposed a game-based argumentation support framework for risk assessment, which in turn is used to translate security goals into security requirements. Similarly, mobile application privacy requirements are dynamic, constantly changing over time and location, and users define them at runtime when interacting with the system. For this purpose, Tun et al.[41] proposed a privacy argument framework that analyzes privacy requirements using extended argumentation language. Furthermore, Jureta et al.[42] proposed the acceptability evaluation framework (ACE), which formally caters to the discussions between stakeholders and requirements engineers on the validity of RE artefacts in the form of a graph. For this purpose, an acceptability condition is defined; if it holds, it means validity is achieved for the given artifact. Elrakaiby et al.[43] proposed the CaRE framework to transform informal, ambiguous, conflicting, and incomplete stakeholder requirements into complete, consistent requirements using abstract argumentation semantics. Similarly, Bagheri et al.[23] proposed an abstract argumentation-based approach to identify and resolve inconsistencies in the requirements specification utilizing Dung's preferred extension. Furthermore, a preferred function is developed that identifies the most inconsistent pairs of requirements. Elrakaiby et al.[22] proposed an argumentation-based method to identify ambiguities during requirements elicitation by interview. Their proposed model is inspired by the ASPIC+ argumentation framework. Similarly, Al-Alshaikh et al.[24] proposed a rationale based approach to define and identify tacit knowledge that is generated during the requirements elicitation process. Their proposed ERBeTK methodology uses the Question Option Criteria (QOC) rationale representation model.

2.3 | Mining Rationale in Software and Requirements Engineering

More recently, software researchers and software tool vendors have started developing different approaches to filter, analyze, validate automatically, and synthesize software artefacts, developer's chat, and user feedback into actionable decisions or suggestions for the software developers. The main focus of these works were to identify useful and instinctive information for software developers, designers, and requirements engineers present in the large amount of user feedback's and software artifacts from different sources, such as: Internet relay chat messages[10], issue tracking system[44], bug reports and design session documents[11], Amazon store[9], and design artifact documents[45,14]. Alkadhi et al.[10] applied different classifiers to automatically capture rationale information in the Internet relay chat messages and classify them into fine-grained rationale elements, i-e, alternative, issues, cons & pro arguments, and decisions. Similarly, Bhat et al.[44] employed distant ML classifiers on the issue tracking system data set to capture and classify architectural design decisions. At the same time, Kurtanovic and Maalaj[9] employed ML algorithms on the Amazon crowd-users reviews to capture rationale concepts and classify them into distant rationale elements, i-e, criteria, issues, decisions, alternatives, and justifications. Wang et al.[8] proposed ArguLens, an argumentation-based automated approach, to analyze and evaluate the developer's opinions and comments about usability issues in the issue tracking systems. They experimented with ML algorithms (SVM and Naïve Bayes) to classify community opinions into various rationale elements. Lopez et al.[46] proposed an approach to capture and retrieve rationale knowledge from the textual documents using different patterns and ontology-based software rationale representations with a stakeholder-in-a-loop for validation purposes. Rogers et al.[11] applied multiple ML algorithms to identify and captured distant rationale elements from the software bug reports and software design session transcripts documents.

2.4 | Comparing with Existing research works

Although, the software literature on acquiring argumentation theory for requirements analysis and validation is rich for in-house software applications. In contrast, little or no research has focused on requirements analysis, conflict negotiation and prioritization using argumentation theory for market-based software applications according to our knowledge to date. Due to the emergence of market-based software apps and a large number of feedback available on the various social media platforms about these apps, it is now pivotal to put in place the CrowdRE approach that emphasizes capturing conflicts and provides negotiation and prioritization remedies using argumentation theory to identify requirements-related information. The following conclusion can be drawn by analysing and validating the argumentation-based requirements acquisition approaches mentioned above:

- For the argumentation-based requirements acquisition, most existing approaches focus on requirements analysis for in-house software applications involving a limited number of end-users and requirements. Thus, lacks identifying conflicts and prioritizing captured requirements-related information for market-based software applications involving many end-users.

  However, Morales-Ramirez and Perini [25] proposed an argumentation-based discussion forum approach to help analysts identify useful information that can be considered software bugs to fix. The approach considers only attacking relationships when identifying requirements-related information. In contrast, issue tracking systems contain supporting and attacking arguments to make informed requirements decisions. Therefore, the proposed CrowdRE-VArg approach effectively identifies requirements-related information by considering their supporting and attacking information to derive the corresponding aggregated strength values.

- In the literature, Kurtanovic and Maalaj [9], Alkadhi et al. [10], and Khan et al. [28] utilize social media platforms such as Amazon reviews, internet relay chat messages, and Reddit forum to extract useful rationale and requirements-related information using ML algorithms. However, their research approaches lack in employing argumentation theory to recover conflicting viewpoints on the run in the end-user conversation and make informed requirements decisions.

- Similarly, Rogers et al. [11] uses existing design documents and bug reports from the issue tracking system to capture rationale elements that are considered pivotal in software evolution and maintenance. Although, their research approach is limited in consideration of more general social media platforms, i.e., Reddit forum, to mine rationale information using argumentation theory to improve software evolution and maintenance further.

- Earlier, Haley et al. [37], Bagheri et al. [23], Khan et al. [26], Ionita et al. [40], and Elrakaiby et al. [43,22] proposed argumentation-based requirements engineering approaches for in-house development. However, Owing to the digital transformation of many industries and society at large (and the subsequent emergence of big data), it has become relevant to consider alternative information sources for requirements, in addition to traditional stakeholders. Unlike, the proposed approach covers the social media platform, i.e., the Reddit forum, to identify conflict-free prioritize requirements-related information for the market-based software maintenance and evolution.

  Furthermore, existing solutions are compared with the proposed CrowdRE-VArg approach by considering various RE activities and proposed research questions to demonstrate its effectiveness and efficiencies in requirements decisions-making, as shown in Table 1

3 | BACKGROUND AND RESEARCH METHODOLOGY

This section describes the study background, research questions and the adopted research methodology to answer these questions.

3.1 | Research Background

Unlike our previously published work, which uses argumentation theory and semantics to identify conflict-free new features or issues together with the winning arguments. This proposed approach uses an abstract valuation argumentation framework to identify aggregated strength values for each new feature or issue by gradually valuating their supporting and attacking arguments. Unlike the previous approach, the proposed approach records requirements-related information for which crowd-users registered many attacking arguments with minimal strength value. Therefore, requirements and software engineers can still access the requirements-related information that poses many attacking arguments in the Reddit forum. We further extended this research work, which is summarized below:

- In agile software development, developers cannot incorporate all the identified new features or issues in the next iteration; instead, only a subset of these needs to be shortlisted based on their importance. In this approach, we are interested in identifying the relative strength of the captured new features or issues by gradually valuating their supporting and attacking

TABLE 1 Comparisons of CrowdRE-VArg with existing approaches

RE Activities	Existing Approaches	Proposed CrowdRE-VArg
Requirements Elicitation	Kurtanovic and Maalaj[9], Alkadhi et al.[10], and Khan et al.[28] proposed ML-based approaches to capture many rationales and requirements-related information for requirements engineers. Such information requires additional processing to make it useful for software engineers.	In contrast, the proposed CrowdRE-VArg approach more effectively identifies requirements-related information in the Reddit forum by filtering conflicting information using valuation-based argumentation theory to make it presentable for the software engineers to improve future versions.
Crowd Requirements Analysis	Although, Haley et al.[37], Bagheri et al.[23], Khan et al.[26], Ionita et al.[40], and Elrakaiby et al.[43,22] used argumentation theory to analyze requirements and remove conflicts. However, their approaches limit either by considering only the attacking relation between the end-user requirements or limited to the in-house software applications or case studies.	On the other hand, the CrowdRE-VArg approach analyses crowd requirements by considering both supporting and attacking relationships to improve the requirements decision-making process. Also, it scales the crowd requirements analysis process by considering end-user conversations in the Reddit forum to identify conflict-free useful requirements-related information.
Requirements Negotiation & Prioritization	According to our knowledge, the research approaches of Khan et al.[3,27], Haley et al.[37], Bagheri et al.[23], Ionita et al.[40], and Elrakaiby et al.[43,22] provides remedies to achieve negotiation between the end-users over the possible set of requirements using argumentation theory. However, their approaches are limited in prioritizing identified requirements information essential for market-based software development where many requirements-related information are captured from the social media platforms.	In comparison, the CrowdRE-VArg approach utilizes valuation-based argumentation to recover conflicting viewpoints on requirements-related information in the Reddit forum and then prioritize them by gradually valuating their corresponding supporting and attacking arguments. The higher the aggregated value of the captured new features or issues, the higher their priority and vice-versa. A case study on it is elaborated in section 6.2.
RQ-1	To our knowledge, none of the existing approaches used valuation-based argumentation approaches to recover the conflicts arrived on end-user requirements for both in-house and market-based software applications to make informed requirements decisions and develop a prioritized requirements list.	We took this opportunity and developed the CrowdRE-VArg approach by adopting the theory of valuation-based argumentation to recover conflicting viewpoints on the run and make informed-requirements decision-making. Also, the proposed method developed a prioritized list of captured new features or issues. The details are depicted in sections 4.2 and 6.2 by analyzing a case study on the Google Maps mobile application from the Reddit forum.
RQ-2 and 3	Khan et al.[3,27] proposed a bipolar argumentation-based approach to identify conflict-free new features and issues in the Reddit forum. However, in the Reddit forum, end-users submit much requirements-related information making it challenging for the software engineers to consider them for the future version of the market-based application. Also, if a single attacking argument appears against the identified new feature or issue, it forces them to be rejected according to the theory of.	In contrast, we the CrowdRE-VArg approach improves the requirements decision-making process by extending the GVBA that effectively identifies an aggregated score associated with each requirements-related information in the Reddit forum by gradually evaluating their pros and cons arguments. Furthermore, it improves the process by keeping the captured requirements-related information at a lower priority for which end-user registered single or many cons arguments instead of ignoring them. Also, we employed various ML feature extraction techniques to improve the classification results.

arguments instead of identifying conflict-free new features or issues only, together with their winning arguments in the user discussion over the Reddit forum.

- To identify conflict-free new features or issues in the online user conversations, we need to define a threshold, i-e, the Base Score score, that will work as a centre point in identifying conflict-free features or issues. For this purpose, we extended the heuristic rules for the proposed CrowdRE-VArg approach.
- Crowd users submit a bulk of user comments in the Reddit forum, which is considered as an important alternative source for requirements elicitation together with the traditional stakeholders. It is time-consuming and difficult to process each user comment in the Reddit forum to identify requirements-related information. For this purpose, we proposed an automated CrowdRE-VArg approach extended from the bipolar gradual valuation argumentation framework[30], QuAD[32], ESAAF[33], and DF-QuAD[29] to reach a rationale requirements decision and prioritize the conflict-free requirements-related information by gradually valuating the relative strength of arguments (their supporting and attacking arguments).
- To improve the performance of the ML classifiers, we additionally introduce certain textual features, i-e, Part-of-speech tagging, and topic modeling specifically the Latent Dirichlet Allocation (LDA) algorithm.
- Since our data set is imbalanced, building a classifier on imbalanced data set forces the ML classifiers to skew towards the majority class while ignoring the minority class. For this purpose, we additionally tested the classifiers with under-sampling (under-sampling, Tomek Link), oversampling (ADASYN), and advanced hybrid (SMOTEENN, SMOTETomek) approaches to improve the performance of ML classifiers.
- To scale and enhance the ML experiments, we additionally experimented with Decision Tree, K-Nearest Neighbors, and Bernoulli Naive Bayes algorithms along with the Naive Bayes, Random Forest, Support Vector Machine, multi-layer Perceptron Classifier, and Logistic Regression to report the results.

3.2 | Research Questions:

For this paper, we propose the following research questions that are answered through the proposed CrowdRE-VArg approach:

RQ1: How to use an argumentation-based valuation approach for requirements decision-making and prioritization by requirements analysts?

RQ2: How to adopt gradual valuation of user arguments in negotiating the conflicts over the new features or issues to reach a settlement?

RQ3: How to tackle the complexity of reaching a settlement automatically between the crowd-users in larger discussion topics using AI?

3.3 | Research Method

The detailed workflow of the processing of end-user comments from the Reddit forum to reach a settlement on requirements decisions for the identified new features, design alternatives, issues, or main discussion topics embedded in the proposed CrowdRE-VArg approach is shown in Figure. 1 . The research methodology is comprised of three main steps. Firstly, we need to process the annotated end-user comments (adopted from previously published work[27,3]) collected from the Reddit forum against the Google Map mobile application to make it readable for the CrowdRE-VArg approach by converting it to an argumentation graph using an argumentation theory. Whereas the captured end-user comments are classified into the different rationale and requirements-related elements such as new features, issues, claim-supporting, claim-attacking, and claim-neutral using the content analysis approach in our previous work[27,3].

For the **RQ-2**, we proposed a CrowdRE-VArg approach by extending GVBA, QuAD, and ESAAF frameworks that analyze the argumentation tree constructed from the crowd-user comments to identify whether the captured new features, design alternatives, issues, or even the main discussion topics are collectively accepted or rejected. Also, using the proposed CrowdRE-VArg framework, we process and analyze each user comment in the argumentation tree to identify the grounded arguments, new features, or issues that are classified as conflict-free by the definition of Dung's theory[21]. For example, analyzing the crowd-users comments by employing the proposed CrowdRE-VArg approach, if the end-user comment in the argumentation tree is identified as a "new feature" or an "issue" and is not further discussed and commented on (no supporting or attacking arguments are registered) by the crowd-users in the Reddit forum. The proposed CrowdRE-VArg approach identifies it as conflict-free or grounded

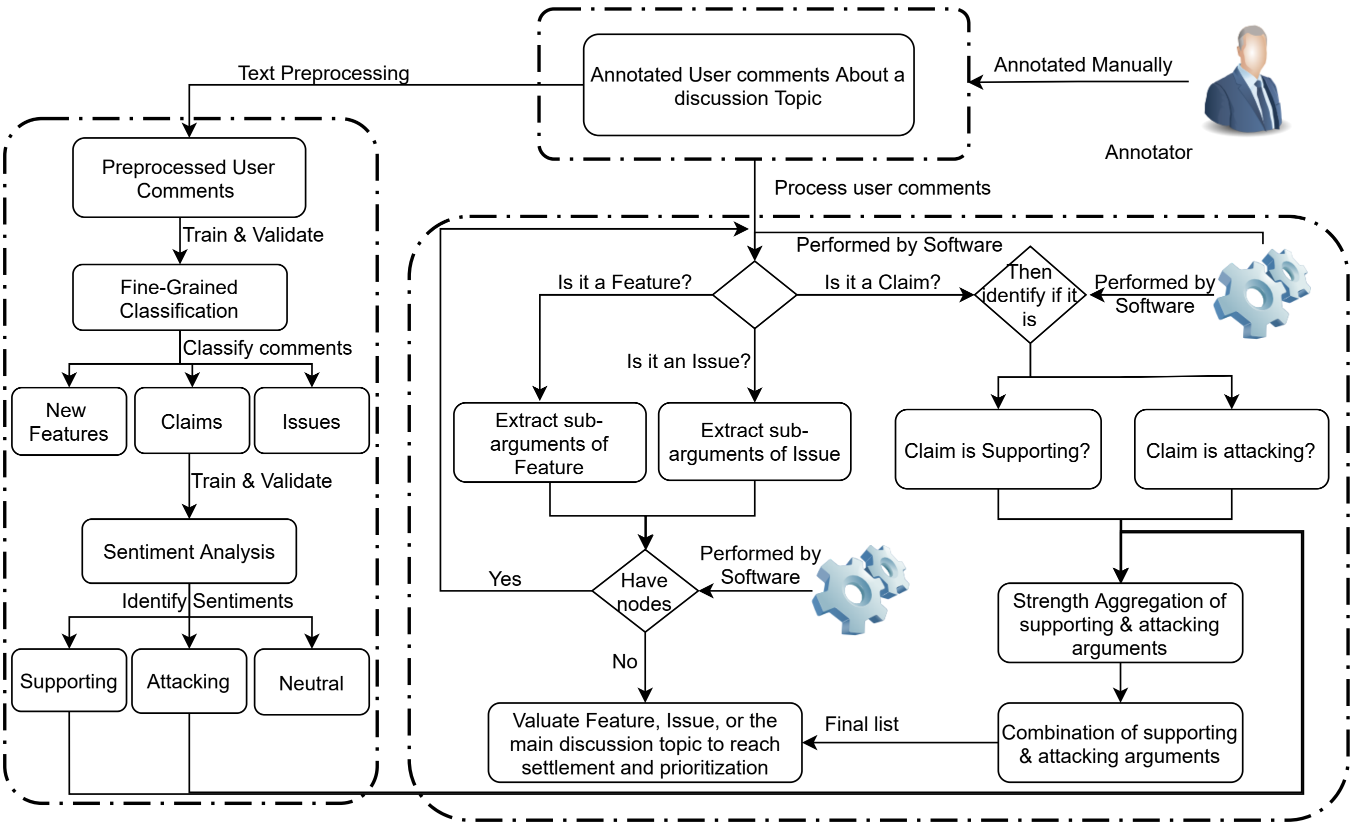

FIGURE 1 Proposed CrowdRE-VArg Approach

using argumentation theory [21]. In contrast, if the proposed CrowdRE-VArg approach identifies that the new feature, issue, or the main discussion topic is further discussed and commented on (provided with supporting or attacking arguments) by the crowd-users, then the CrowdRE-VAg approach triggers the proposed arguments valuation process by identifying the strength values of the new features, issues, or the main discussion topics using their corresponding supporting or attacking arguments. Whereas the proposed CrowdRE-VArg approach utilizes the strength aggregation and combination function adopted from the DF-QuAD [29] framework together with the base score to identify the strength values of each identified new feature, issue, or the main discussion topic. The base score is considered pivotal in requirements decision-making, particularly in requirements prioritization and negotiations. As a result, the proposed CrowdRE-VArg approach captured a conflict-free set of new features, issues, and main discussion topics prioritized by the strength values derived using the concepts of aggregation and combination function. Each argument in the argumentation tree is processed automatically by the software/algorithm proposed to support the automation of the CrowdRE-VArg approach. The proposed CrowdRE-VArg is elaborated in detailed in section 4.1.

For **RQ-3**, in order to scale the proposed CrowdRE-VArg approach by catering to many crowd-users feedback from the Reddit forums; we employ different ML algorithms to classify the crowd-users comments into the different rationale and requirements-related information. For this purpose, we first classify user comments into fine-grained rationale elements, such as new features, issues, and claims. Secondly, we performed the sentiment analysis experiment by classifying only those user comments identified as claim rationale elements into supporting, attacking, and neutral claim elements. Finally, after classifying the end-user comments into different rationale elements, we identify the strengths of the captured new features, issues, or main discussion topics by gradually valuating their supporting and attacking arguments using the proposed CrowdRE-VArg approach.

3.4 | Research Data set

To illustrate and experiment with the proposed CrowdRE-VArg approach, we have curated the end-user comments data set from the Reddit forum. The forum is an online popular social media platform on which registered users submit content such as textual

TABLE 2 Summary of the sample data sets

ID.	Discussion Topic	# of comments
1	Google Maps is testing a combined commute to replace driving and transit	72/72
2	Google Maps Introduces Location Sharing	701/781
3	Take control of your commute with Google Maps	401/435
4	I wish Google Maps had an option to give me less verbose directions, but not entirely mute.	467/552
5	Google Maps Will Soon Allow You to Share Your ETA, Add a Shortcut to Routine Destinations, and Create a Map of Your Location History	583/803
6	Google Maps Now Warns You If You Are Navigating To A Place That Will Be Closed When You Arrive	511/560
7	Google Maps for android is finally rolling out multi-waypoint directions	388/445

posts, links, and images. These posts are organized by various subjects, which are written in the Reddit forum by user communities and are referred to as "subreddits" such as news, technology, android, science, video games, etc. Other registered users that are known as "Redditors" can join subreddits of their interests. The Redditors can submit distant posts in the joined Subreddits and gets upvotes or downvotes from the other registered Redditors. There are only 10 million and 2.3 million registered users in Technology and Android subreddits, respectively. The Reddit forum is gaining enormous popularity amongst the social media platforms; as of October 2019, there are 430 million active registered members with 199 million active posts or subreddits as compared to the 330 million registered users in April 2018 with 130 million active posts[47].

To analyze and experiment with the proposed CrowdRE-VArg approach, we utilized the same data set used in previously published work[27,3] but with different intentions and goals. For example, we are interested in identifying the strength values of each captured new feature or issue in the user conversation by gradually evaluating their supporting and attacking arguments, if there are any. It results in a prioritized list of identified new features or issues that can be utilized by the software developers for the next release of the software application under discussion to improve its performance and functionalities. The data set includes a user comment associated with an end-user name, the number of up and down votes received by that comment, the time stamp, a unique comment id, and a child id. The main discussion topics included in the data set possess the child id of "0", as it is the beginning of the discussion. For this research article, seven discussion topics from the Reddit forum cover various aspects of the Google Maps mobile application, such as demanding a new feature, various issues about its interface, and user experiences, as shown in Table. 2 . A link [i] is provided to access the actual discussion topics in the Reddit forum. The objectives of choosing the discussion topics are: the discussion topic must be relevant to the Google Maps mobile application; it must be either discussing a new feature introduced, an issue encountered, end-user experience with it, and the size of user comments against the main discussion topic must be greater than fifty containing some useful requirements-related information and possess argumentation structure. We collected 3123 crowd-user comments from the seven online discussion topics in the Reddit forum about the Google Maps mobile application. Where the first discussion topic, as shown in Table. 2 is used as a proof-of-concept example for the proposed CrowdRE-VArg approach. In contrast, the remaining six discussion topics (3046 user comments) are used as an experimental data set for conducting fine-grained ML and sentiment analysis experiments.

3.5 | Argumentation-Tailored User comments & Annotation to run CrowdRE-VArg approach

In this sub-section, we overview the rationale elements captured from the end-user conversations in the Reddit forum and the annotation process that labels raw user comments in the data set to make them parsable for the CrowdRE-VArg approach. Identifying requirements-related information from the Reddit forum using the proposed CrowdRE-VArg requires an annotated data set as an input, which is achieved by intertwining the following two steps with the proposed approach. These steps are explained in detail in the previously published work[27,3]. However, we briefly introduce both of these steps for better readability.

[i]https://drive.google.com/file/d/1WPv-MurlZ7lH-j2T3eBRC64PWM9I5yAw/view?usp=sharing

3.5.1 | Capturing User Statements

In our earlier work[27,3], we did a comprehensive analysis of the crowd-user comments in the data set to identify rationale and requirements-related information using the grounded theory approach of Strass and Corbin[48]. We were interested in designing an approach that captures those users statements in the Reddit forum that contains rationale and requirements-related information. For example, user statements that demand or request a new feature, a user preference on the application interface design, an issue faced by the users, or feedback about user experience with the software application. The following rationale and requirements concepts are identified due to the detailed analysis of user comments: new feature or design alternative, issue, claim-attacking, claim-supporting, and claim-neutral. These elements are also elaborated in the coding guide used by the coders to annotate the user comments in the data set. The definition of each identified rational element and their examples are shown in Table. 3 . Additionally, during the manual analysis of the user comments, feedback in the user discussions is organized in a hierarchy. Their immediate child comments are linked to their parent's comments. Therefore, we captured and identified rationale associations between the rationale elements, as shown in Table. 4 . To run the proposed CrowdRE-VArg approach, we need to replicate these steps. Therefore, we borrowed the existing analysis procedure of the crowd-users comments for the proposed CrowdRE-VArg approach to design a research approach that captures and identifies requirements-related information's strength values by gradually evaluating their supporting and attacking arguments.

TABLE 3 Types of User Statements

ID.	Example User Statements from Reddit	Label	Definitions of the Label
1	"All I want Google Maps to do is to let me download the entire map for my country and store it offline."	Feature	Crowd-users demand or suggest a new feature for the software application under discussion or suggest an alternative solution
2	"This is really cool, thanks"	Claim-supporting	The crowd-users registers a claim in support of an issue, new feature, or the main discussion topic
3	"Too bad the store hours on Google Maps are inaccurate "	Claim-attacking	The crowd-users registers a claim to oppose an issue, new feature, or the main discussion topic
4	"Do Americans usually eat a combination of sugar eggs and cheese for breakfast?, All i know is the Canadian sausage end egg mc muffin is pretty damn good "	Claim-neutral	The crowd-users registers a claim that neither oppose or support an issue, new feature, or the main discussion topic
3	"Did they remove this from Hangouts?", "It has a cycle option?", "But will assistant be able to do it via voice command?"	Issue	Problems or concerns of crowd-users that require further elaboration, clarification, and discussion. It also presents a problem with the application.

3.5.2 | Labeling User Comments

Currently, the crowd-users comments collected from the Reddit forum are in the raw form. To make it parsable for the CrowdRE-VArg approach, it needs to annotate the end-user comment manually by the coders using the developed coding guidelines. For this purpose, we adopted the same annotation steps from the previously published work performed in the following steps. To highlight, first, we rearrange the user comments in the data set manually according to their original sequence in the Reddit forum to preserve their order. Secondly, a coding form in the Excel environment was created to annotate the user comments. The user comments in the data set were chosen as the primary information to be annotated by the coder to enable a more fine-grained content analysis. Thirdly, the actual peer-review task was performed to annotate the user comments. Each coder received the coding guideline and the coding form containing a complete list of the users comments against the main discussion topics. Two coders took part in the annotation process. In the next step, the annotation results were merged from the two coders and accessed their reliability by calculating the inter-coder agreement and Cohen's kappa[49]. Furthermore, to minimize the disagreement and conflict between the two coders, a coding guideline [ii] was adopted that includes rationale definitions along with their detailed examples for the different rationale concepts. The average time to complete the reviews coding task was 14 working hours. The standard inter-coded agreement for annotating the user comments was calculated by combining the coding results from the two coders. The average inter-coder agreement between the two coders was 90%, while the Cohen's kappa for the two coders was 67%, which is a substantial agreement on the Cohen's kappa scale. A sample of the labelled data is shown in Table. 5 .

TABLE 4 Types of relations between Arguments

Relation Type	Relation Symbol	Relation Definition
Supports	$\rightsquigarrow$	When a crowd-user argument supports another argument
Attacks	$\longrightarrow$	When a crowd-user argument attacks another argument
alternatives	$\dashrightarrow$	When a crowd-user suggests a design alternative or a new feature
issue	$\not\longrightarrow$	When a crowd-user submits a problem or challenge to the software application under discussion

4 | VALUATION-BASED REQUIREMENTS ARGUMENTATION: PRIMITIVES AND POSTULATES

In Reddit forum, crowd-users submit their opinions about a recently released new feature, a change made in the interface, or an issue triggered while using the current software. We parse this user feedback and identify helpful and valuable information for requirement engineers and software developers. To reach a rationale-based requirements decision about a new feature, design alternative, issues, or the main discussion topic, one needs to gradually valuate the relative strength of their supporting or attacking arguments in the Reddit forum. For example, the different solutions or alternatives for an issue are gradually valuated based on their supporting and attacking claims to identify the best possible solution for the said issue[32]. For this purpose, argumentation theory[21] provides a built-in reasoning mechanism, which foresees human debate and reasoning processes, helps in constructing arguments, comparing them, evaluating them in some respect, and finally, concludes to find out whether any of them are rebutted[50,51].

By making rationale-based decision-making, employing the theory and semantics of the valuation argumentation framework[31] on the identified rationale elements and their relationships can help capture conflicts between the gathered crowd-users feedback from the Reddit forum. This gradual valuation of the arguments' strength would help resolve run-time conflicts and achieve negotiations between a large pool of crowd-users over the captured new features, design alternatives, or issues. In turn, the valuation process helps accommodate multiple viewpoints and lack of consensus between the crowd-users, arriving at the design interface's variability and adaptivity. Requirements argumentation can be employed as a phase where multiple viewpoints are captured and rationalized to achieve consensus between the stakeholders over much requirements-related information. It helps requirement analysts and software developers identify prioritized and negotiated requirements-related information using a valuation-based argumentation framework. The detailed steps of the proposed CrowdRE-VArg approach are discussed below:

4.1 | Crowd-based Requirements Engineering by Arguments Valuation (CrowdRE-VArg)

We proposed the CrowdRE-VArg approach that analyzes crowd-users comments in the Reddit forum. The proposed approach assists requirements analysts in negotiating the conflicts over the captured new features, design alternatives, or the issues between the diverse stakeholders on the run by finding a settlement that would satisfy the involved crowd-users in the ongoing user discussion while keeping the rationale. The proposed CrowdRE-VArg approach is extended from the GVAAF[31], GVBA[30], QuAD[32], ESAAF[33], and DF-QuAD[29]. Whereas,

[ii] https://drive.google.com/open?id=19VF4YjH6BumZm_mJnvE0WWz4slXGPklv

TABLE 5 Sample of the coding form

Users_Comments	Comment_Id	Parent_Id	User_Votes	Rationale_Type	Claim_Type
Google Maps is testing a combined Commute tab to replace Driving and Transit	9clgh7	0	401	claim	discussion topic
Maps is getting way too bloated and uses way too much rich-media. The explore tab could be its own app	e5bpnup	9clgh7	37	claim	attacking
I don't know how practical it would be especially for people that don't drive or the others that don't take public transits	e5bh1by	9clgh7	125	claim	attacking
Can't they just rename the tab dynamically based on whether you selected Drive or Public Transport in the "how do you get to work?" question they ask you? This is how Google Assistant handles things.	e5bi1ce	9clgh7	54	new feature	

GVAAF: is extended from Dung's AAF, which gradually valuates an argument's strength based on its attacking argumentsi. Suppose $\mathcal{W}$ be an AAF, $\mathcal{A}$ represents new features, issues, or arguments, and "Atk" shows the attacking relation between them. Also, supposed $\mathcal{W}$ be a completely ordered set of new features and issues that contain a maximum and minimum element ($\mathcal{V}_{\mathcal{MAX}}$, $\mathcal{V}_{\mathcal{MIN}}$), respectively. The strength values for a new feature, issue or the main discussion topic are in the range or interval of [0,1][31]. Then, the argument strength valuation is a function defined as $\mathfrak{v} : \mathcal{A} \rightarrow \mathcal{V}$, such that:

(I). $\forall\, b \in \mathcal{A}$, $\mathfrak{v}(b) \geq \mathcal{V}_{\mathcal{MIN}}$

The strength of any new feature or issue identified in the Reddit forums must be greater or equal to $\mathcal{V}_{\mathcal{MIN}}$ that is derived by gradually valuating the strength of their attacking arguments. In that case, the strength value of any new feature or issue identified must be greater or equal to "0".

(II). $\forall\, b \in \mathcal{A}$, if $\mathcal{R}^{-}(b) = \emptyset$, then $\mathfrak{v}(\mathfrak{b}) = \mathcal{V}_{\mathcal{MAX}}$

While the strength of any new feature or issue identified in the Reddit forums must be equal to $\mathcal{V}_{\mathcal{MAX}}$ if other users in the Reddit forum have not registered any attacking arguments in response. In example 1, the strength values of (a_1) and (a_3) will be "1" as the crowd-users did not registered any attacking arguments against (a_1) and (a_3).

(III). $\forall\, b \in \mathcal{A}$, if $\mathcal{R}^{-}(b) = \{b_1, b_2, \ldots\ldots, b_n\} \neq \emptyset$, then $\mathfrak{v}(\mathfrak{b}) = \mathcal{G}(\mathcal{H}(\ \mathfrak{v}(b_1), \mathfrak{v}(b_2), \ldots\ldots, \mathfrak{v}(b_n)))$

Furthermore, suppose other end-users register attacking arguments in response to the captured new features, issues, or the main discussion topics. In that case, their strength is identified with the composite function ($\mathcal{G}(\mathcal{H}(\ \mathfrak{v}(b_1), \mathfrak{v}(b_2), \ldots\ldots, \mathfrak{v}(b_n))))$ by gradually valuating the strength of their attacking arguments. Firstly, the composition function aggregates the values of the direct attackers of the captured new features, issues, or the main discussion topic. At the same time, the attackers take their defenders into account. Secondly, it synthesizes the effect of the "direct attack" on the strength of the captured requirements-related information, such as the strength of the new feature, issue or the main discussion topic decreases if the strength of the "direct attack" increases and vice-versa.

Example 1: The Figure. 2 , represent the Dung's AAF containing only attacking ($\longleftrightarrow$) relationship between the arguments. The direct attackers of an argument (b) are the elements of the set $\mathcal{R}^{-}(b)$. In example 1, the direct attackers of an argument (a_0) are "a_1" and "a_2". At the same time, the direct defenders of an argument (a_0) are direct attackers of the elements of a set $\mathcal{R}^{-}(b)$. In example 1, the direct defender of the root argument (a_0) in "(a_3)". To identify the strength value of the new feature, issue, or an argument (a_0) using the definition of GVBA. For this purpose, the strength values of the arguments (a_3) and (a_1) is equal to 1, while the value of argument (a_2) is identified by computing ($\mathcal{H}(\ \mathfrak{v}(1/1 + a_3))$= (1/1+1)= 0.5. Similarly we can identify the strength value of (a_0) = 0.4 by valuating the attacking arguments using the composite function ($\mathcal{G}(\mathcal{H}(\ \mathfrak{v}(a_1), \mathfrak{v}(a_2))))$.

GVBA: Unlike GVAAF, the GVBA framework handles and models two independent relations between the arguments, such as supporting and attacking relationships. Such a framework suits the Reddit forum, where end-users registers both supporting and attacking arguments in response to the new features, issues, or the main discussion topics[28]. The supporting relation carries

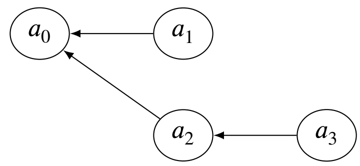

FIGURE 2 Abstract Argumentation Framework

positive information and vice-versa. In the GVBA framework, the final strength of requirements-related information is identified by gradually valuating the strength of their supporting and attacking arguments. The GVBA framework is define as: Suppose, $\mathcal{V}$ be a completely ordered set containing maximum element ($\mathcal{V}_{\mathcal{MAX}}$) and minimum element ($\mathcal{V}_{\mathcal{MIN}}$). Also, ($\mathcal{H}_{Supp}$) and ($\mathcal{H}_{Atk}$) be two ordered sets of supporting and attacking arguments, wheres $\mathcal{V}^*$ be the set of finite sequences of the elements of $\mathcal{V}$.

Now, suppose $\langle$ $\mathcal{A}$, R_{atk}, R_{supp} $\rangle$ be a bipolar argumentation framework ()[34]. Then, the gradual local valuation on $\langle$ $\mathcal{A}$, R_{atk}, R_{supp} $\rangle$ is a function $\mathfrak{v} : \mathcal{A} \rightarrow \mathcal{V}$, such that:

- $\forall$ $c \in \mathcal{A}$, if, $R^{-}_{atk}(c) = \{a_1,a_2,.....,a_n\} \neq \emptyset$ and $R^{-}_{supp}(c)= \{s_1,s_2,.....,s_n\} \neq \emptyset$, then, $\mathfrak{v}(\mathfrak{c}) = \mathcal{G}(\mathcal{H}_{supp}(\ \mathfrak{v}(s_1),\mathfrak{v}(s_2),....,\mathfrak{v}(s_n))$ $\mathcal{H}_{atk}(\mathfrak{v}(a_1),\mathfrak{v}(a_2),....,\mathfrak{v}(a_n)))$
- Where the functions, $\mathcal{H}_{supp}$: $\mathcal{V}^{-} \rightarrow \mathcal{W}_{supp}$ and $\mathcal{H}_{atk}$: $\mathcal{V}^{-} \rightarrow \mathcal{W}_{atk}$, are used to valuate the quality of the supporting and attacking arguments that arrives against the captured the new feature, issue, or the main discussion topics in the Reddit forum.

ESAA: The GVAAF and GVBA frameworks consider only supporting and attacking arguments when evaluating the requirements-related information in the Reddit forum. In contrast, the ESAA framework[33] extended from social abstract argumentation (SAA)[52], associates votes along their supporting or attacking arguments when evaluating requirements-related information. It is defined as: An ESAA framework is a 4-tuple $\langle\mathcal{A}$, R_{atk}, R_{supp},$\mathcal{V}\rangle$, where $\langle\mathcal{A}$, R_{atk}, $R_{supp}\rangle$ is an BAP, while $\mathcal{V}$: $\mathcal{A} \rightarrow \mathcal{N} \times \mathcal{N}$ is a function that maps arguments, new features, or issues in the Reddit forum to the number of their negative and positive votes.

Furthermore, to evaluate the arguments in ESAA framework, Evripidou and Toni[33] proposed semantics that handle both supporting and attacking relationships between the arguments. The semantic framework is a 7-tuple $\mathcal{S} = \langle\mathcal{M}, \pm,\wedge,\vee,\perp,\oplus, \uplus\rangle$. The total mapping function $\mathcal{N} : \mathcal{A} \rightarrow \mathcal{M}$ is a social argumentation model under semantics $\mathcal{S}$. Therefore, given an ESAA framework $\langle\mathcal{A}$, R_{atk}, R_{supp},$\mathcal{V}\rangle$ and the semantic $\mathcal{S}$, then the valuation of an argument $b \in \mathcal{A}$ is:

$\mathcal{N}^{+}(b) = (\pm(b) \wedge \perp \vee \{\mathcal{N}^{+}b_i: (b_i,b) \in R_{atk}\}\ ) \uplus (\pm(b) \oplus \vee \{\mathcal{N}^{+}b_i: (b_i,b) \in R_{supp}\}\ )$

Where, the $\mathcal{N}^{+}(b)$ is refer as the valuation of an argument b in $\mathcal{N}$, while the other operators are defined as:

$\mathcal{M}$ is a totally ordered set that contains all possible valuation of the requirements-related information with top element($\mathcal{M}_{Max}$) and bottom element($\mathcal{M}_{Min}$). $\pm$: $\mathcal{N} \times \mathcal{N} \rightarrow \mathcal{M}$ is a function that valuates the social support for every argument b based on its accumulated negative and positive votes. It also assigns a base score $\pm(b)$ to b. The operator $\wedge$: $\mathcal{M} \times \mathcal{M} \rightarrow \mathcal{M}$ defines an algebraic operation on the argument, new feature, or issue valuation in the social argumentation model to identify their strength by gradually valuating their supporting and attacking arguments. The operator ($\wedge$) is used to aggregate the individually valuated attacking and supporting arguments of b in the social argumentation model ($\mathcal{N}$). The operators $\vee$: $\mathcal{M} \times \mathcal{M} \rightarrow \mathcal{M}$, $\perp$: $\mathcal{M} \rightarrow \mathcal{M}$, and $\oplus$: $\mathcal{M} \times \mathcal{M} \rightarrow \mathcal{M}$ define binary and unary algebraic functions used to synthesize the effects of the attacks and supports coming towards the argument b. In the social argumentation model ($\mathcal{N}$), the aggregated valuation of the supporters and attackers are combined together with the $\pm(b)$ utilizing the $\vee$,$\perp$, and $\oplus$ operators. Finally, the operator $\uplus$: $\mathcal{M} \times \mathcal{M} \rightarrow \mathcal{M}$ define a binary algebraic function used to maps the supporting and attacking pair of values obtained in the previous step into a single valuation of b in the social argumentation model.

QuAD: Baroni et al.[32] proposed the QuAD framework to automatically evaluate the positions or solutions put upfront in the debates in an issue-based information system (IBIS) by gradually valuating their supporting and attacking argument. The QuAD framework is inspired by the ESAAF framework by improving the arguments valuation process for selecting best suited alternative solution in an IBIS. It is defined as:

Definition 5: The QuAD framework is a 5-tuple $\langle\mathcal{A}, Atk, Supp,\mathcal{R}, \mathcal{BS}\rangle$, such that:

- $\mathcal{A}$ is a defined as finite set of solutions, answers, or arguments in an argumentation framework.
- *Atk* is a finite set of attacking arguments that arrived in response to the proposed solution or answer.

- *Supp* is a finite set of supporting arguments that arrived in response to the proposed solution or answer.
- $\mathcal{R} \subseteq (Atk \cup Supp) \times (\mathcal{A} \cup Atk \cup Supp)$ is an acyclic binary relation in QuAD framework.
- $\mathcal{BS}$: $(\mathcal{A} \cup Atk \cup Supp) \rightarrow \mathcal{I}$ is total score function for the various alternatives proposed in response to the issue in QuAD framework, where $\mathcal{BS}(b)$ is the base score of $b \in \mathcal{A}$.

  The sets $\mathcal{A}$, *Atk*, and *Supp* are pairwise disjoint. Furthermore, Baroni et al.[32] examined and evaluated various design scenarios with the relevant experts, where they found that $\mathcal{BS}$ is considered an important factor when analyzing the different alternatives for a solution based on their positive and negative arguments. In the QuAD framework, the strength of arguments is determined from their $\mathcal{BS}$ and the aggregated strength of their supporters and attackers in an argumentation tree. The functions $\mathcal{F}_{atk}$: $\mathcal{I} \times \mathcal{I} \rightarrow \mathcal{I}$ and $\mathcal{F}_{supp}$: $\mathcal{I} \times \mathcal{I} \rightarrow \mathcal{I}$ for this purpose are defined as, for $\mathfrak{s}_0$, $\mathfrak{s} \in \mathcal{I}$:

  $\mathcal{F}_{atk}(\mathfrak{s}_0, \mathfrak{s}) = \mathfrak{s}_0$ - $\mathfrak{s}_0$. $\mathfrak{s} = \mathfrak{s}_0$. (1 - $\mathfrak{s}$) and $\mathcal{F}_{supp}(\mathfrak{s}_0, \mathfrak{s}) = \mathfrak{s}_0 + (1 - \mathfrak{s}_0)$. $\mathfrak{s} = \mathfrak{s}_0 + \mathfrak{s}$ - $\mathfrak{s}_0$. $\mathfrak{s}$

  The above equations elaborate on how the strength of a generic argument $\mathfrak{s}_0$ is obtained by gradually modifying its base score ($\mathfrak{s}_0$) in the presence of a single supporter or attacker, respectively having strength ($\mathfrak{s}$).

Additionally, Rago et al.[29] proposed the DF-QuAD approach by extending the QuAD framework, which foresees certain discontinuities that might arise when identifying the strength of an argument. For example, when many arguments in an argumentation tree attack an argument with no support. It might bring the score of that argument close to zero, yet adding single support to that argument will increase its strength score to a comparatively high value and vice versa. The DF-QuAD framework resolves these discontinuities by defining a single argument's strength aggregation function, such as, for a given $\mathcal{S}$ = ($\mathfrak{s}_1$,$\mathfrak{s}_2$,.....,$\mathfrak{s}_k$), where $k \geq 1$, then

- if $k = 0 : \mathcal{E}(\mathcal{S}) = 0$; if $k = 1 : \mathcal{E}(\mathcal{S}) = \mathfrak{s}_1$; if $k = 2 : \mathcal{E}(\mathcal{S}) = \mathcal{F}(\mathfrak{s}_1, \mathfrak{s}_2)$; or if $k > 2 : \mathcal{E}(\mathcal{S}) = \mathcal{F}(\mathcal{E}(\mathfrak{s}_1,\mathfrak{s}_2,....,\mathfrak{s}_{k-1}), \mathfrak{s}_n)$, where
- $\mathcal{F}(\mathfrak{s}_1, \mathfrak{s}_2) = \mathfrak{s}_1 + (1 - \mathfrak{s}_1)$. $\mathfrak{s} = \mathfrak{s}_1 + \mathfrak{s}$ - $\mathfrak{s}_1$. $\mathfrak{s}$

CrowdRE-VArg: Above, we elaborated various argumentation frameworks that inspire the proposed CrowdRE-VArg approach. The underlying principle of the CrowdRE-VArg approach is to reach a rationale-based requirements decision-making by identifying the strength of the arguments, new features, issues, or the main discussion topics by gradually valuating the strength of their supporting and attacking arguments in the argumentation tree constructed from the crowd-users conversation in the Reddit forums. The proposed CrowdRE-VArg approach is deeply inspired by the QuAD framework with different intentions and goals. Unlike the CrowdRE-VArg approach, emphasis on the end-user discussion in the Reddit forum to identify and capture pivotal conflict-free requirements-related information. Furthermore, we developed and extended heuristic rules to identify conflict-free prioritized new features or issues in the Reddit forum. Formally, the CrowdRE-VArg can be defined as:

Definition 1: The CrowdRE-VArg framework is a 7-tuple $\langle \mathcal{T}, \mathcal{F}, \mathcal{I}\ \mathcal{A}_{Atk}, \mathcal{A}_{Supp}, \mathcal{R}, \mathcal{BS} \rangle$, such that:

- $\mathcal{T}$ is a finite set of online discussion topics in the Reddit forum, where diverse crowd-users discuss software new features, hot issues, or alternative solutions in an argumentative fashion.
- $\mathcal{F}$ is defined as a finite set of new features or design alternatives identified in the online discussion topics ($\mathcal{T}$).
- $\mathcal{I}$ is defined as a finite set of captured issues or problems faced by the crowd-users while using the software application under discussion in the online discussion topics ($\mathcal{T}$).
- $\mathcal{A}_{Atk}$ is defined as a finite set of captured attacking arguments in the online Reddit forum that arrived in response to the main discussion topic, new feature, alternative design, or an issue.
- $\mathcal{A}_{Supp}$ is defined as a finite set of identified supporting arguments in the online Reddit forum that arrived in response to the main discussion topic, new feature, alternative design, or an issue.
- $\mathcal{R} \subseteq (\mathcal{F} \cup \mathcal{I} \cup \mathcal{T}) \times (\mathcal{A}_{Atk} \cup \mathcal{A}_{Supp})$ is an undirected binary relation between the elements of the CrowdRE-VArg framework that results in an argumentation tree. Where $\mathcal{F}$, $\mathcal{I}$, and $\mathcal{T}$ are the identified new features, issues, or main discussion topics in the Reddit forum. On the other hand, $\mathcal{A}_{Atk}$ and $\mathcal{A}_{Supp}$ are their corresponding supporting and attacking arguments. The elements ($\mathcal{F}$, $\mathcal{I}$, and $\mathcal{T}$) are represented with $\cup$ to include all the requirements-related information in the Reddit forum. Similarly, these requirements-related elements have supporting, attacking, or both arguments supporting requirements decision-making, are presented with $\cup$. Furthermore, below, we formally defined each CrowdRE-VArg element that results in $\mathcal{R}$ and how they are presented in an argumentation tree constructed from end-user comments.

- $\forall$ (a_j,a_k) $\in \mathcal{A}_{Atk}$, such that, j≠k and j,k $\neq 0$, then a_j attacks a_k, is represented as $a_j \rightarrow a_k$ in an argumentation tree constructed from the crowd-users arguments in the Reddit forums.
- $\forall$ (a_j,a_k) $\in \mathcal{A}_{Supp}$, such that, j≠k and j,k $\neq 0$, then a_j supports a_k, is represented as $a_j \rightsquigarrow a_k$ in an argumentation tree constructed from the crowd-users arguments in the Reddit forum.
- $\forall$ (a_j,a_k) $\in \mathcal{F}$,such that, j≠k and j,k $\neq 0$, then a_j is a new feature proposed by crowd-users in response of a_k in the Reddit forum is represented as $a_j \dashrightarrow a_k$ in an argumentation tree constructed from the crowd-users arguments.
- $\forall$ (a_j,a_k) $\in \mathcal{I}$,such that, j≠k and j,k $\neq 0$, then a_j is an issue or challenge proposed by crowd-users in response of a_k in the Reddit forum is represented as $a_j \nrightarrow a_k$ in an argumentation tree.
- $\exists$ $a_j \in \mathcal{T}$,such that, $a_{j=0}$, then a_j is the main discussion topic represented as a root node in the argumentation tree constructed from the crowd-user arguments in the Reddit forum.

- $\mathcal{BS}$: $(\mathcal{F} \cup \mathcal{I} \cup \mathcal{T} \cup \mathcal{A}_{Atk} \cup \mathcal{A}_{Supp}) \rightarrow \mathcal{I}$ is a total strength score function for the new features, issues, or the main discussion topics identified by the CrowdRE-VArg framework, where $\mathcal{BS}(c)$ is the base score of $c \in$ either $(\mathcal{F}, \mathcal{I}, \mathcal{T}, \mathcal{A}_{Atk}, \mathcal{A}_{Supp})$.

  In Literature, the $\mathcal{BS}$ is considered pivotal in requirements decision-making, particularly, in requirements prioritization and negotiation[53,54,55]. Also, Cocarascu and Toni[20], Leite and Martins[52], Evripidou and Toni[33], Carstens and Toni[56,57], and Baroni et al.[58] highlighted its importance in identifying the strength an argument in an argumentation tree. Therefore, $\mathcal{BS}$ is considered when identifying the strengths of captured new features, issues, or the main discussion topics by the CrowdRE-VArg approach.

  The strength of an argument, new feature, issue, or the main discussion topic in an argumentation tree is identified recursively by gradually valuating their supporting and attacking arguments. For this purpose, we adopted the strength aggregation ($\mathcal{S}_{Agr}$) and Strength combination ($\mathcal{S}_{Com}$) functions from Rago et al.[29], which are:

- The $\mathcal{S}_{Agr}$[20] function, given **k** arguments(supporting/attacking) having strength values ($\mathfrak{s}_1$,$\mathfrak{s}_2$,....,$\mathfrak{s}_k$) is given as:

$$\mathcal{S}_{Agr}(\mathfrak{s}_1, \mathfrak{s}_2, ..., \mathfrak{s}_k) = \begin{cases} 0 & \text{k} = 0 \\ 1 - \log \prod_{j=1}^{k}(|1 - \mathfrak{s}_j|) & \text{k} > 0 \end{cases} \quad (1)$$

  In the CrowdRE-VArg approach, the strength of each argument, new feature, issue, or the main discussion topic is calculated by recursively valuating the strength of their supporting and attacking arguments, respectively. Each argument, new feature, issue, or the main discussion topic is considered as a sub-root node in the tree if they have child nodes (supporting or attacking) to identify their aggregated supporting and attacking strength based on their $\mathcal{BS}$, respectively.

- The $\mathcal{S}_{Com}$[20], for an argument, new feature, issue, or the main discussion topic with a $\mathcal{BS}$ $\mathfrak{s}_0$, their corresponding attacking arguments having strength ($\mathfrak{s}_1$,$\mathfrak{s}_2$,....,$\mathfrak{s}_k$) and their supporting arguments having strength ($\mathfrak{s}'_1$,$\mathfrak{s}'_2$,....,$\mathfrak{s}'_k$) is elaborated as, for $\mathfrak{s}_a = \mathcal{S}_{Agr}(\mathfrak{s}_1,\mathfrak{s}_2,....,\mathfrak{s}_k)$ and $\mathfrak{s}_s = \mathcal{S}_{Agr}(\mathfrak{s}'_1,\mathfrak{s}'_2,....,\mathfrak{s}'_k)$:

$$\mathcal{S}_{Com}(\mathfrak{s}_0, \mathfrak{s}_a, \mathfrak{s}_s) = \begin{cases} \mathfrak{s}_0 & if \mathfrak{s}_a = \mathfrak{s}_s \\ \mathfrak{s}_0 - \log(\mathfrak{s}_0 * |\mathfrak{s}_s - \mathfrak{s}_a|) & if \mathfrak{s}_a > \mathfrak{s}_s \\ \mathfrak{s}_0 + \log(1 - \mathfrak{s}_0) * |\mathfrak{s}_s - \mathfrak{s}_a|) & if \mathfrak{s}_a < \mathfrak{s}_s \end{cases} \quad (2)$$

  The $\mathcal{S}_{Com}$ is used to identify the dialectical strength of an argument, new feature, issue, or the main discussion topic captured in the crowd-user conversation, once their supporting and attacking arguments strength have been aggregated using $\mathcal{S}_{Agg}$ function. The intuition behind the $\mathcal{S}_{Com}$ is that if the arguments attacking strength against an argument, new feature, issue, or the main discussion topic is smaller (respectively, larger), then the $\mathcal{BS}$ is increased (respectively, decreased) by utilizing the absolute value of the difference between the supporting and attacking strength to obtain a strength score which is likely closer to 1 (respectively, 0).

- Finally, to calculate the strength value of a new feature, issue, or the main discussion topic with a $\mathcal{BS} = \mathfrak{s}_0$, k attacking arguments having strength $\mathfrak{s}_a = \mathfrak{s}_1$,$\mathfrak{s}_2$,....,$\mathfrak{s}_k$, and l supporters arguments of strength $\mathfrak{s}_s = \mathfrak{s}'_1$,$\mathfrak{s}'_2$,....,$\mathfrak{s}'_l$, is defined as:

$$Strength\_Score(b) = Str(\mathcal{S}_{Com}(\mathfrak{s}_0, \mathcal{S}_{Agr}(\mathfrak{s}_1, \mathfrak{s}_2, ..., \mathfrak{s}_k), \mathcal{S}_{Agr}(\mathfrak{s}'_1, \mathfrak{s}'_2, ..., \mathfrak{s}'_l))) \quad (3)$$

Where $b \in \mathcal{F}$, $\mathcal{I}$, or $\mathcal{T}$. By this function, we identify the strength of new features, issues, or the main discussion topics identified and captured in the user conversation. The function is adopted from the Cocarascu & Toni[20] approach. The strength values are calculated by valuating their supporting and attacking arguments along with their $\mathcal{BS}$. It helps requirements analysts in decision-making by identifying the most important, negotiated, and prioritized new features and issues rationale elements while keeping the rationale behind for the next release of the software application under discussion.

Additionally, we adopted heuristic rules from our previous approach[27], which are further modified to identify prioritized conflict-free new features and issues in the Reddit forum for which other users did not register supporting or attacking arguments.

Rule No 1. If $(a_i,a_j) \in R_{alt}$, a_j is a new feature, then $set_{alt}= \{a_i\} \cup \{a_j\}$. It is considered conflict-free when an argument is identified as a new feature, and there is no attacking argument. It is by definition of requirements-related arguments and argumentation theory.

Rule No 2. If $(a_i,a_j) \in R_{iss}$, a_j is an issue or a challenge, then $set_{iss}= \{a_i\} \cup \{a_j\}$. Also, it is considered conflict-free when an argument is identified as an issue, and there is no attacking argument. It is by the definition of requirements-related arguments and argumentation theory.

Rule No 3. If the identified new features or issues did not receive any attacking or supporting arguments, their strength would be considered as the default $\mathcal{BS}$ value. It is by definition of requirements-related arguments and argumentation theory (the grounded arguments are always considered conflict-free and are included in the conflict-free arguments set).

Rule No 4. If the strength value of the identified new features or issues is equal to or greater than the default $\mathcal{BS}$ value, it is considered winning requirements-related information. By the definition of argumentation theory and requirements-related arguments, many supporting arguments come in response to the captured new features or issues.

Rule No 5. If the strength value of the identified new features or issues is less than the default $\mathcal{BS}$ value. In that case, it is either rejected or kept at a low priority due to the possible conflict between the end-user comments in the Reddit forum. It is by definition of argumentation theory and requirements-related arguments, where many attacking arguments arrive in response to the captured new features or issues.

Below, we explained how the extended form (CrowdRE-VArg) of the QuAD and DF-QuAD frameworks is employed in RE and forum context with an actual crowd-user conversation snippet taken from the Reddit forum.

4.2 | Running Example to demonstrate the CrowdRE-VArg Approach (A Case Study)

To reach a rationale requirements decision, we selected a crowd-users discussion snippet from topic 1 (shown in Table 2) to demonstrate the proposed CrowdRE-VArg approach. We choose similar discussion topics to elaborate and evaluate the proposed CrowdRE-VArg approach that has been used in our past work[27]. The crowd-users discussion thread is shown in Figure. 3 , in which crowd-user demands a new feature (denoted by "a_0") to change the design interface of the Google Map mobile application, against which other end-users give supporting and attacking arguments, hence results in argumentation. In the previous studies, it has been identified that crowd-users often suggest new features and report hot issues faced by them while using the software application in the Reddit forum[28,3]. To make the CrowdRE-VArg approach effective, we need to convert the end-user conversation snippet, as shown in Figure. 3 to an argumentation tree, as depicted in Figure. 4 , to make the data parsable for the CrowdRE-VArg approach. For this purpose, in our previous research work, we proposed an algorithm[27], which automatically converts user discussion in the Reddit forum to an argumentation tree. To better understand the proposed CrowdRE-VArg, we manually converted the crowd-user conversation into an argumentation tree by assigning "Tag" to each user comment, as shown in the Figure. 3 . In the argumentation graph of crowd-user comment, the symbol ($\rightsquigarrow$) represents a supporting relation, the symbol ($\rightarrow$) shows an attacking relation, and the symbol ($\dashrightarrow$) shows a new feature relation between the corresponding user comments, as shown in the Figure. 4 .

To support the CrowdRE-VArg approach, we developed and designed certain rules to help negotiate the conflicts over the new features or issues between the different crowd-users on the run by finding a settlement that would satisfy the involved end-users in the ongoing discussion in the Reddit forum. For this purpose, to support the rationale-based requirements decision-making, we need a certain method, which will assign a quantitative valuation, referred to as a strength function, that answers each node in the argumentation tree. The strength function for any user comment (b) in the argumentation tree is defined as,

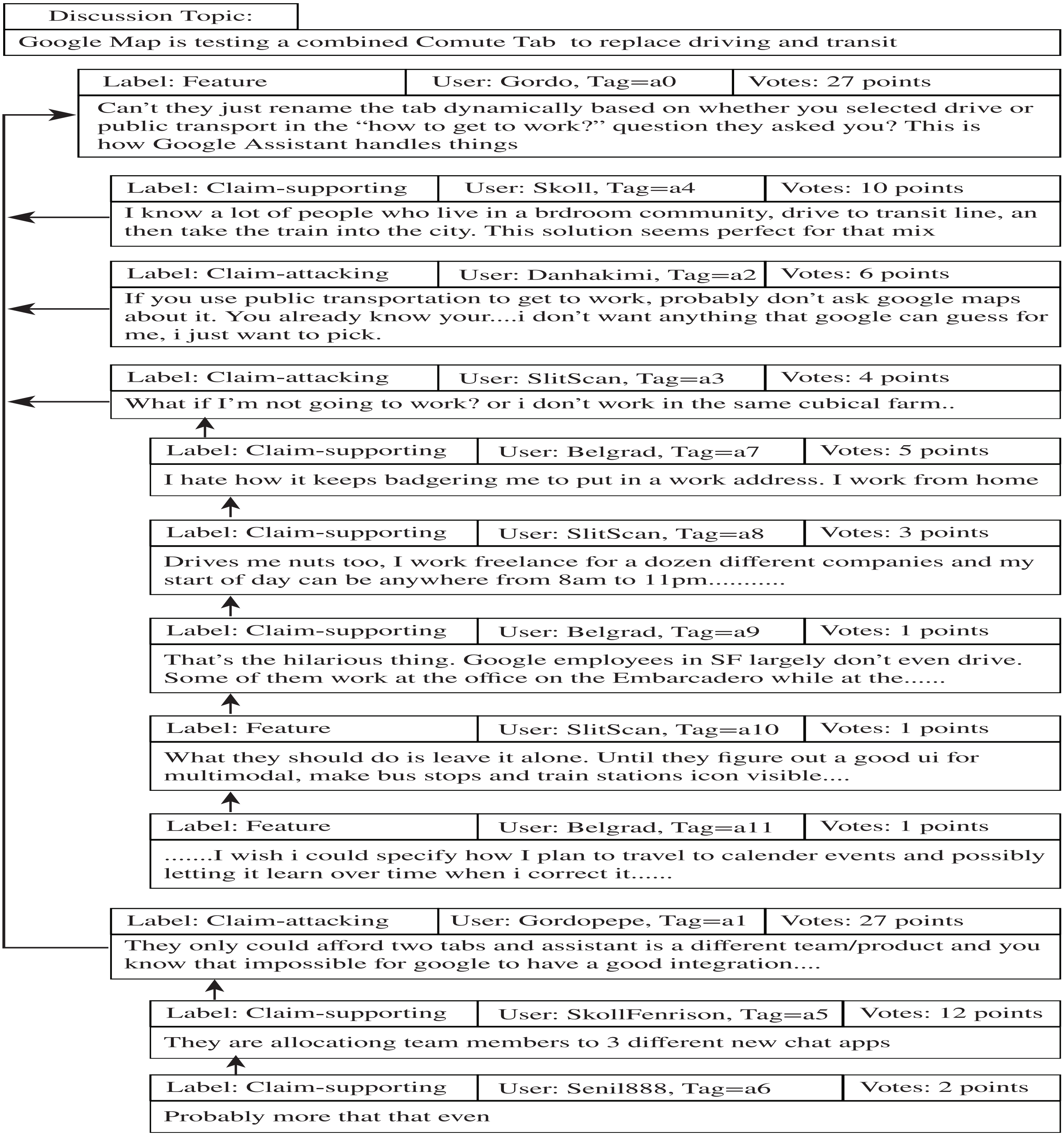

FIGURE 3 Users' discussion about a design alternative

Strength_Score(b) = Str($\mathcal{S}_{Com}$($\hat{s}_0$, $\mathcal{S}_{Com}(\hat{s}_1,\hat{s}_2,....,\hat{s}_k)$, $\mathcal{S}_{Com}$ $(\hat{s}'_1,\hat{s}'_2,....,\hat{s}'_l)$)), which is explained in detailed in section 4.1. The final strength of the user comment, new feature, or issue in the argumentation tree depends on its $\mathcal{BS}$ and the final strength of its attacking and supporting arguments. Therefore, for the proposed CrowdRE-VArg approach, we adopted a bottom-up approach to combine these three elements and make rationale-based requirements decisions. We assume the $\mathcal{BS}$ for all the user comments $b \in (\mathcal{F} \cup \mathcal{I} \cup \mathcal{T} \cup \mathcal{A}_{Atk} \cup \mathcal{A}_{Supp})$, $(\mathcal{BS}(b) = 0.5$. We will use the same ($\mathcal{BS}$) score for all the arguments, new features, and issues in the argumentation tree constructed from the user comments in the Reddit forum.

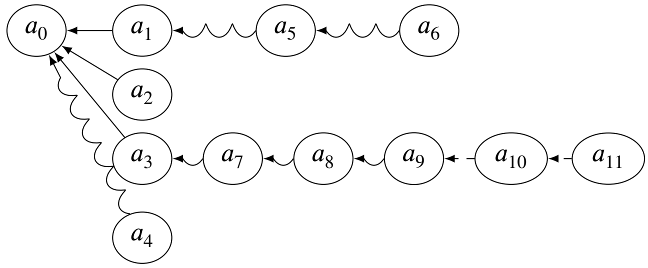

FIGURE 4 Bipolar Argumentation tree constructed from the users discussion in Figure 3

Thus, processing the argumentation tree using the CrowdRE-VArg framework, at the bottom, two new features are identified, having labels "a_{11}" and "a_{10}", as shown in Figure 4 . Therefore, according to Rule NO. 1 of the CrowdRE-VArg approach, these new features are grouped as they are connected and conflict-free. Also, according to Rule No. 3, the strength of these identified new features or design alternatives is the default value of the $\mathcal{BS}$, as they don't have any attacking and supporting arguments. Furthermore, to identify the strength of the new feature "a_0" using equation 3, we gradually valuated its supporting and attacking arguments by following a bottom-up approach. For this purpose, we first need to identify an argument strength aggregation and combination, as explained in equation 1 and 2, respectively with $\mathcal{BS}$ (a) = $\mathfrak{s}_0$. For example, using equation 1 and 2, we identify the strength of an argument "a_8" in the bipolar argumentation tree, as shown in Figure 4 , to calculate the strength of a new feature "a_0", as follow:

Strength_Score(a_8) = Str($\mathcal{S}_{Com}$(0.5, $\mathcal{S}_{Agr}$(), $\mathcal{S}_{Agr}$ ($\mathfrak{s}_9$))), in this case, we don't have attacking arguments, therefore, $\mathcal{S}_{Agr}$() is represented with no arguments. Now computing this, using equation 1 and 2, we get the strength value of $a_8 = 0.75$. By similar way, we can identify the strength value of the new feature a_0 proposed by crowd-users against the main discussion topic in the Reddit forum, such as:

Strength_Score(a_0) = Str($\mathcal{S}_{Com}$(0.5, $\mathcal{S}_{Agr}$($\mathfrak{s}_1$,$\mathfrak{s}_2$,$\mathfrak{s}_3$), $\mathcal{S}_{Agr}$ ($\mathfrak{s}_4$))), computing the values we get, Strength_Score(a_0) = Str($\mathcal{S}_{Com}$(0.5, $\mathcal{S}_{Agr}$(0.875,0.5,0.937), $\mathcal{S}_{Agr}$ (0.5))). Now, using the aggregation function, shown in equation 1, we identify the strength of a_0 supporting and attacking arguments, respectively. By doing so we get, $\mathfrak{s}_a$= 1- (1-0.875)(1-0.5)(1-0.937)= 0.996 and $\mathfrak{s}_s$= 1-(1-0.5)= 0.5. Next, using combination function, shown in equation 2, we can identify the strength value of the new feature "a_0", based on the values of $\mathfrak{s}_a$ and $\mathfrak{s}_s$. In this case, $\mathfrak{s}_a > \mathfrak{s}_s$, as 0.996 > 0.5, therefore, Strength_Score(a_0) = 0.5 - (0.5*(|0.5-0.996|)) = **0.252**. The proposed CrowdRE-VArg framework identifies the strength value of a new feature or design alternative "a_0" proposed against the main discussion topic in the Reddit forum by gradually valuating its supporting and attacking arguments while keeping the rationale behind. The new feature with the strength value **0.252** may be discarded or kept at a low priority (Rule No. 5). The CrowdRE-VArg framework gives weightage to each crowd-user feedback in the argumentation tree when performing argumentation-based requirements decision-making, which improves end-user confidence in the software application under discussion. Additionally, it makes the requirements decision-making process transparent by documenting its rationale and providing a prioritized list of new features, design alternatives, and issues rationale elements. Similarly, the CrowdRE-VArg framework calculates an accumulated strength value of the main discussion topic (root node) by gradually valuating its supporting and attacking arguments and all other remaining alternative features or issues in response to the main discussion in the Reddit forum. Finally, we get a prioritized list of the designed alternatives, new features, or hot issues with quantified strength values. The strength values of the identified new features, design alternatives, or issues close to 1 have been given high priority and vice-versa. It will help requirements analysts and software developers by giving more importance and preference to the high-priority new features or issues in the next release of the software application.

5 | AUTOMATICALLY CLASSIFYING USER COMMENTS

On social media platforms, end-users submit the bulk of user comments, such as on Twitter; end-users send around 500M messages in a single day[59]. At the same time, in the Reddit forum, there are 199 million active crowd-users post and 430 million active users as of 2019[47]. Also, Pagano et al.[60] found in an empirical research study that mobile apps in play stores received approximately 23 user reviews per day while the other popular mobile apps, such as Facebook and Twitter, etc. receive on average 4275 user reviews per day. Previously, to elaborate the proposed CrowdRE-VArg approach with the actual crowd-user conversation in the Reddit forum, we selected a discussion topic (“Google Maps is testing a combined commute to replace driving and transit”) shown in Table. 2 that contains only 72 user comments, and it’s easy to manage it manually. This example also serves as a proof-of-concept case for the proposed CrowdRE-VArg approach. In contrast, most discussion topics in the Reddit forum contain many end-users comments, which in-turns become difficult, and challenging to manage manually. In particular, for the proposed CrowdRE-VArg approach, we combined six similar discussion topics on the Google Maps mobile application, resulting in 3046 user comments. Therefore, we require various ML algorithms to monitor their performance in automatically classifying crowd-users feedback into different rationale and requirements elements that are identified previously in section 3.5. For this purpose, we conducted two experiments; firstly, the end-user statements in the data set are classified into new features, claims, and issue rationale elements. Secondly, a sentiment analysis approach is applied to the user comments identified as claim-rational elements to capture the crowd-user opinions, i.e., supporting, attacking, or neutral. We require to do so, as the CrowdRE-VArg approach operates on label data set to reach a settlement and negotiate the conflicts arrived on the captured new feature, issues, or the main discussion topics. The detailed process of ML experiments is shown in Figure. 1 .

5.1 | Classifying Crowd-Users Feedback into Claims, Issues, or New Features

This experiment employs different classifiers to automatically capture and classify end-users feedback into the various rationale and requirements elements, i-e, new features, issues, and claims. In our previous published work [28], it is identified that the claim rationale element occurs in a high percentage (62%) as compared to other rationale elements. Therefore, we decided first to classify the end-user feedback into new features, issues, and claim rationale elements. Secondly, we consider only those user comments that will be identified as the claim rationale element. The details are given below:

5.1.1 | Setup

To run the first experiment of fine-grained classification, we select eight ML algorithms to compare their performance in classifying end-user feedback into the different rationale and requirements elements. The classifiers selected for this purpose are Logistic Regression (LR), Random Forests (RF), Multilayer Perceptron Classifier (MLP), Support Vector Machine (SVM), Naive Bayes Multinomial(MNB), Decision Tree (DT), K-Nearest Neighbors (KNN), and Bernoulli Naive Bayes (BNB). We selected these algorithms for the ML classification due to their good performance and popularity on textual data sets [61,62,63,64].

5.1.2 | Text Pre-Processing

The text preprocessing step of the CrowdRE-VArg approach takes single end-user feedback in the data set as an input string. We performed extensive NLP activities in the text preprocessing step to clean the data set and possibly improve the performance of the ML classifiers. For this purpose, we first eliminate the HTML tags in the end-user feedback if there are any. Then, it removes the URLs present inside the end-user comments. Next, in the preprocessing pipeline, three basic filtering steps are applied to the end-user comments, such as 1) eliminating all types of brackets if there are any, 2) eliminating special symbols and characters, and 3) removing bad characters in the end-user comments. Next, we replaced the auxiliary verbs and other frequently used abbreviations in the data set with their full form. For this purpose, we created a python dictionary, where the auxiliary verbs and abbreviations are put as a key in the dictionary. In contrast, their corresponding full form is put as a value in the python dictionary. For example, the auxiliary verb “isnt” is replaced with “is not”, “afaik” is replaced with “as far as i know”, etc. Next, we converted the end-user comments text into lower case characters and divided each user’s comments into tokens. Finally, a lemmatization technique of NLP is applied to the end-user comments to reduce the words to their root. Its purpose is to normalize the text in the user comments of the data set. Furthermore, we decided not to remove stop words from the user comments in the data set. While manually analyzing and annotating the user comments, some stop words are frequently used to represent end-user intentions. For example, the stop words “would” and “will” were used as a possible indicator for new feature or design alternatives, the stop words “does”, “isn’t”, and “did” were used as an indicator to represent issues in the data set.

5.1.3 | Features Engineering

We identified potential features for the ML classifiers by reviewing relevant literature on Reddit forum comments, app store feedback, and tweet analytics [18,61]. We adopted Bag of words (BOW) and Term Frequency-Inverse Document Frequency (TF_IDF) textual features for each ML algorithm widely used textual features identified in the literature with better performance [62]. In the BOW approach, a dictionary of all the words in the corpus is created. For each user comment in the data set, the frequency of each word is counted. The (TF_IDF) approach works on the principle that high-frequency words in the corpus may not provide much high information gain. In contrast, rare words in the corpus may contribute more weight to the classifiers. For this purpose, we utilize CountVectorizer and TfidfVectorizer methods of scikit-learn [iii] python library for BOW and (TF_IDF), respectively. Secondly, we adopted the N-gram, the most widely used textual feature that groups together a continuous sequence of N-tokens in a user comment. We adopted “ngram_range” parameter of TfidfVectorizer having values (1,3) as the minimum and maximum length of the words in the corpus. We apply these settings with each classifier to capture and identify the common patterns of words that appear together in the corpus, which can be a possible indicator of the presence of intentional elements in the end-user comments. For example, “wish to”, and “I want a” phrases might be considered as possible indicators for a new feature element in the data set. Wheres, the phrases “Is this”, and “not sure how” are considered as the possible indicators for an issue rationale element in the data set. We also checked the performance of each ML classifier by using both “char” and “word” analyzer together with the ngram of the TfidfVectorizer method.

In literature, Part-of-speech (POS) tagging has been used as a possible textual feature of the classification algorithms that can enhance the classifier performance[65]. For example, end-users usually use past tense verbs in their comments to report a previous user experience, specifically a past problem or issue with the software application. Whereas end-users typically use a verb with future tense to elaborate on a hypothetical situation, i-e, suggest a new feature or design alternatives. For this purpose, we tag the type of verbs in each user’s comments using the TextBlob library of python, which is used to perform common NLP tasks such as POS tagging, sentiment analysis, classification, etc. The verbs used in the experiment are MD (Model auxiliary, i-e, may, should), VB (base form, i-e, think), VBZ (3rd person, present, i-e, he thinks), VBP (1st person, present, i-e, I think), VBD (past tense, i-e, they thought), VBN (past participle, i-e, A sunken ship), and VBG (present participle, i-e, thinking is fun). When training the ML algorithms, the presence or absence of each verb tag in the user comment was used as a text feature.

Besides the verb tags, we trialed other POS taggers as text features in the ML classifiers. It was identified by manually analyzing the POS tags in the user comments that not all tags have a meaningful correlation to the different end-user comment classes. In contrast, the tags CD (cardinal digit), MNP, and WRB (wh-adverb, i-e, how) are identified to have a meaningful correlation to the various rationale classes. Additionally, we employ topic modeling as a textual feature with the different ML classifiers to find the hidden semantic structure in the user comments. We apply the Latent Dirichlet Allocation (LDA)[66] algorithm of topic modeling, which has been widely adopted in software and requirements engineering to probabilistically process a large amount of crowd-user data in the social media platforms and cluster similar groups of requirements-related information together in a supervised manner[2]. For this purpose, we used the powerful “gensim” python library to create the bigram representation of the crowd-users reviews. To train the LDA algorithm, we need to provide a control variable to display several topics in the textual corpus, referred to as “K”. In our case, we choose the value of "K=20". This topic model vector will be used as a textual feature for every user comment with the different ML algorithms to identify the performance of the classifiers in identifying rationale elements in the user comments. To make the LDA algorithm perform better, we remove the most common (stop words) and rarest words in the textual corpus.

5.1.4 | Data Imbalance

Since we are developing multiple classifiers with several classification labels, and for any single classification class, the number of crowd-user comments in that class (minority class) is outnumbered by the crowd-users comments in all the other classification classes (majority class). It can force the ML classifier to become biased, resulting in most, if not all, of the crowd-user comments being identified or classified as the majority class. Also, the imbalanced data set for the text classification problem is considered one of the technical issues and challenges in supervised ML.[67].

The developed research data set of end-users comment is also quite imbalanced, as, while labeling the crowd-user comments, the majority (68%) of user’s comments were identified as “claims” rationale element, 22% of user comments were identified as new features, and 10% of crowd-users comments were identified as issues. The ML classifiers can be artificially re-balanced by either under-sampling the majority class or oversampling the minority class to resolve this issue. In the over-sampling approach, additional sentences are added to the training sample by automatically generating minority sentences or re-sampling the existing sentences in the minority class. While in the under-sampling approach, some of the sentences in the majority class are discarded to balance the data across different classification labels.

For the CrowdRE-VArg approach, we experimented with both under-sampling and over-sampling techniques to identify the best methodology to re-balance the data and optimize the performance of ML algorithms in classifying user comments into

[iii] https://scikit-learn.org/stable/accessed on 22-04-2021

TABLE 6 Most accurate classifiers to mine user comments from Reddit forum

Rationale Codes	Classifier and Features Used	Data Balancing	P	R	F1
Claims	LR + Countvectorizer + POS + verbs	SMOTE	0.752	0.600	0.666
	MNB + TF-IDF + POS + verbs + analyzer = ’word’	SMOTE	**0.845**	0.462	0.596
	MLP + TF-IDF + POS + verbs + analyzer = ’word’	R. Under Sampler	0.805	0.555	0.656
	MLP + TF-IDF + POS + verbs + analyzer = ’word’	SMOTE	0.782	**0.782**	**0.753**
	SVM + TF-IDF + POS + analyzer = ’word’ + ngrams	R. Under Sampler	0.753	0.730	0.735
	LR + TF-IDF + POS + analyzer = ’word’ + ngrams	SMOTE	0.789	0.730	**0.757**
	MLP + TF-IDF + POS + analyzer = ’word’ + ngrams	SMOTE	0.781	**0.751**	**0.764**
	LR + TF-IDF + POS + analyzer = ’char’ + ngrams	R. Under Sampler	**0.823**	0.580	0.679
	DT + TF-IDF + POS + analyzer = ’char’ + ngrams	SMOTE	0.748	0.685	0.715
	MLP + TF-IDF + POS + analyzer = ’char’ + ngrams	SMOTE	0.806	0.712	0.755
	MLP + Countvectorizer + POS + verbs	ADASYN	0.747	0.671	0.704
	LR + TF-IDF + POS + analyzer = ’char’ + ngrams	SMOTETomek	0.807	0.719	**0.760**
	MLP + Scaler + POS + ngram + LDA Topic modeling	SMOTETomek	0.727	0.458	0.559
	LR + TF-IDF + POS + analyzer = ’word’ + ngrams	ADASYN	0.800	0.711	**0.751**
	LR + TF-IDF + POS + analyzer = ’char’ + ngrams	ADASYN	0.808	0.703	**0.751**
	LR + TF-IDF + POS + analyzer = ’char’ + ngrams	SMOTEENN	**0.824**	0.499	0.620
New Features	LR + Countvectorizer + POS + verbs	SMOTE	0.444	0.540	0.481
	MNB + TF-IDF + POS + verbs + analyzer = ’word’	SMOTE	0.393	**0.735**	0.507
	MLP + TF-IDF + POS + verbs + analyzer = ’word’	R. Under Sampler	0.475	0.526	0.494
	MLP + TF-IDF + POS + verbs + analyzer = ’word’	SMOTE	**0.527**	0.575	**0.540**
	SVM + TF-IDF + POS + analyzer = ’word’ + ngrams	R. Under Sampler	0.439	0.451	0.435
	LR + TF-IDF + POS + analyzer = ’word’ + ngrams	SMOTE	**0.524**	0.587	**0.546**
	MLP + TF-IDF + POS + analyzer = ’word’ + ngrams	SMOTE	**0.528**	0.573	**0.540**
	LR + TF-IDF + POS + analyzer = ’char’ + ngrams	R. Under Sampler	0.420	**0.605**	0.491
	DT + TF-IDF + POS + analyzer = ’char’ + ngrams	SMOTE	0.390	0.454	0.416
	MLP + TF-IDF + POS + analyzer = ’char’ + ngrams	SMOTE	0.493	**0.603**	0.538
	MLP + Countvectorizer + POS + verbs	ADASYN	0.474	0.529	0.491
	LR + TF-IDF + POS + analyzer = ’char’ + ngrams	SMOTETomek	0.497	0.592	0.536
	MLP + Scaler + POS + ngram + LDA Topic modeling	SMOTETomek	0.337	0.455	0.382
	LR + TF-IDF + POS + analyzer = ’word’ + ngrams	ADASYN	0.516	**0.603**	**0.549**
	LR + TF-IDF + POS + analyzer = ’char’ + ngrams	ADASYN	0.486	0.592	0.530
	LR + TF-IDF + POS + analyzer = ’char’ + ngrams	SMOTEENN	0.393	**0.683**	0.495
Issue	LR + Countvectorizer + POS + verbs	SMOTE	0.246	0.408	0.301
	MNB + TF-IDF + POS + verbs + analyzer = ’word’	SMOTE	0.249	0.415	0.308
	MLP + TF-IDF + POS + verbs + analyzer = ’word’	R. Under Sampler	0.237	0.584	0.333
	MLP + TF-IDF + POS + verbs + analyzer = ’word’	SMOTE	0.332	0.380	0.347
	SVM + TF-IDF + POS + analyzer = ’word’ + ngrams	R. Under Sampler	0.389	0.400	0.377
	LR + TF-IDF + POS + analyzer = ’word’ + ngrams	SMOTE	0.375	0.433	0.394
	MLP + TF-IDF + POS + analyzer = ’word’ + ngrams	SMOTE	0.384	0.390	0.380
	LR + TF-IDF + POS + analyzer = ’char’ + ngrams	R. Under Sampler	0.415	**0.733**	**0.526**
	DT + TF-IDF + POS + analyzer = ’char’ + ngrams	SMOTE	0.454	0.539	0.490
	MLP + TF-IDF + POS + analyzer = ’char’ + ngrams	SMOTE	**0.483**	0.583	**0.523**
	MLP + Countvectorizer + POS + verbs	ADASYN	0.266	0.344	0.293
	LR + TF-IDF + POS + analyzer = ’char’ + ngrams	SMOTETomek	**0.492**	**0.607**	**0.539**
	MLP + Scaler + POS + ngram + LDA Topic modeling	SMOTETomek	0.123	0.304	0.174
	LR + TF-IDF + POS + analyzer = ’word’ + ngrams	ADASYN	0.372	0.460	0.402
	LR + TF-IDF + POS + analyzer = ’char’ + ngrams	ADASYN	**0.481**	0.592	**0.530**
	LR + TF-IDF + POS + analyzer = ’char’ + ngrams	SMOTEENN	0.389	**0.665**	0.487

different rationale elements. For this purpose, we adopted the existing state-of-the-art approaches to re-balance the data by either adding or removing the samples, i-e SMOTE (Synthetic Minority Over-sampling Technique)[68], ADASYN[69] for over-sampling approaches, while Random Under-sampling[70] and Tomek Link[71] techniques are used for under-sampling. Furthermore, we also employed advanced data re-balancing techniques by combining under-sampling and over-sampling techniques, i-e, SMOTEENN (SMOTE of over-sampling and Edited Nearest Neighbor (ENN) of under-sampling)[70,72] and SMOTETomek[72].

5.1.5 | Training and Evaluation of the Machine Learning Algorithms

To train and validate the ML classifiers, we applied the standard 10-fold cross-validation techniques over the crowd-users data set containing the user conversation about the Google Maps mobile application (last six discussion topics) in the Reddit forum shown in Table. 2 , where 9 folds of the cross-validation are used to train the ML algorithms while the remaining 1-fold is used for the validation purpose of the classifiers. The training and testing process is repeated 10 times by rotating the testing and training folds. The evaluation for the ML classifiers is computed and reported by calculating the average results among the 10 runs of cross-validation. To evaluate the performance of the ML classifiers, we employ the standard evaluation metrics, i.e., Precision (P), Recall (R), and F-measure (F1). The P and R for the ML classifiers are calculated as:

$$P_k = \frac{TP_k}{TP_k + FP_k} \qquad R_k = \frac{TP_k}{TP_k + FN_k}$$

Where TP_k is the number of end-users feedback that is correctly classified as the type of k, FP_k is the number of end-users feedback that is wrongly classified as the type of k. At the same time, FN_k is the number of end-users feedback wrongly classified as not the type of k where F1 is the harmonic mean of the P and R.

5.1.6 | Results for Fine-Grained Classification

Table 6 shows the average 10-fold cross-validation results for various ML algorithms to classify users' comments in the Reddit forum into claims, new features, and issues rationale elements. The classification results are presented for each rationale element by combining textual features with the corresponding ML algorithm that gives the best results. The experimental results show that ML classifiers using over-sampling re-balancing approaches outer-performs under-sampling re-balancing approaches in most cases. The possible reason might be the loss of important information in discarding samples when performing under-sampling. Furthermore, at times the hybrid re-balancing techniques (SMOTETomek and SMOTEENN) performed better with the LR classifier, particularly when classifying "issue" rationale element.

Overall, the **LR** and **MLP** classifiers outer-perform the other ML classifiers by achieving better recall, precision, and f-measure in classifying user comments into the claims, new features, and issue rationale elements. Both **MLP** and **LR** classifiers gives the best average f-measure scores for classifying claim rationale element, i.e., **0.764** and **0.760** respectively. Similarly, Both **MLP** and **LR** classifiers gives the best average f-measure scores for classifying new feature and issue rationale elements using the over-sampling re-balancing approach, i.e., **0.540**, **0.549** for new feature and **0.523**, **0.539** for issue rationale element respectively. Furthermore, we didn't get better results by employing LDA topic modeling as a textual feature with the different ML classifiers to find the hidden semantic structure in the user comments. One reason might be the data sparsity, as the crowd-user conversations in the Reddit forum are quite diverse; the topics selected by the LDA algorithm from the textual corpus do not represent the meaningful information for the rationale elements. To reach the root cause of low performance, such as low precision, etc., we need to introduce other evaluation metrics, i.e., confusion metric, Receiver operating characteristic (ROC), Area Under Curve (ROC), and mean absolute error where the main purpose is to minimize the False Negative and False Positive. However, we showed a classification result for each rationale element by employing an ML algorithm with LDA topic modeling textual feature in Table 6 , whereas the results obtained are comparatively degraded to the results obtained with other ML algorithms having various textual features. Based on the experimental results as shown in Table 6 , we can choose either **LR** or **MLP** classifier as the best algorithm to classify user comments in the Reddit forum into claims, new features, or issues rationale elements. The best textual features with the corresponding ML algorithm that results in the best classification results are also shown in Table 6 .

5.2 | Sentiment Analysis

In the first experiment of the CrowdRE-VArg approach, we classify user comments in the Reddit forum into claims, new features, and issues rationale elements. In contrast, the proposed CrowdRE-VArg approach identifies conflict-free new features or issues

TABLE 7 Most accurate classifiers to identify sentiments user comments in the Reddit forum

Rationale Codes	Classifier and Features Used	Data Balancing	P	R	F1
Claims-attacking	LR + Countvectorizer + POS + verbs	SMOTE	0.353	0.452	0.393
	LR + TF-IDF + POS + verbs + analyzer = 'word'	SMOTE	0.427	0.494	0.450
	LR + TF-IDF + POS + verbs + analyzer = 'word'	R. Under Sampler	0.397	**0.575**	**0.464**
	LR + TF-IDF + POS + verbs + analyzer = 'word'	SMOTE	0.419	0.482	0.444
	MLP + TF-IDF + POS + analyzer = 'word' + ngrams	R. Under Sampler	0.393	**0.608**	**0.473**
	LR + TF-IDF + POS+ verbs + analyzer = 'word'	TomekLinks	**0.474**	0.308	0.367
	LR + TF-IDF + POS + analyzer = 'char' + ngrams	R. Under Sampler	0.380	0.557	0.445
	LR + TF-IDF + POS + analyzer = 'char' + ngrams	SMOTE	0.391	0.449	0.410
	MLP + TF-IDF + POS + analyzer = 'char' + ngrams	SMOTE	0.401	0.423	0.403
	LR + TF-IDF + POS + verbs + analyzer = 'word'	R. Over Sampler	0.435	0.498	0.456
	LR + TF-IDF + POS + verbs + analyzer = 'word'	SMOTETomek	0.424	0.482	0.444
	RF + Scaler + POS + ngram + LDA Topic modeling	SMOTETomek	0.345	0.392	0.361
	LR + TF-IDF + POS + analyzer = 'word' + ngrams	R. Over Sampler	0.429	0.471	0.444
	SVM + TF-IDF + POS + analyzer = 'char' + ngrams	R. Over Sampler	0.447	0.248	0.310
	MLP + TF-IDF + POS + verbs + analyzer = 'word'	TomekLinks	**0.452**	0.326	0.368
Claim-Neutral	LR + Countvectorizer + POS + verbs	SMOTE	0.501	0.389	0.434
	LR + TF-IDF + POS + verbs + analyzer = 'word'	SMOTE	0.558	0.534	0.539
	LR + TF-IDF + POS + verbs + analyzer = 'word'	R. Under Sampler	**0.560**	0.488	0.515
	LR + TF-IDF + POS + verbs + analyzer = 'word'	SMOTE	0.547	0.527	0.532
	MLP + TF-IDF + POS + analyzer = 'word' + ngrams	R. Under Sampler	0.528	0.435	0.472
	LR + TF-IDF + POS+ verbs + analyzer = 'word'	TomekLinks	0.531	**0.616**	**0.561**
	LR + TF-IDF + POS + analyzer = 'char' + ngrams	R. Under Sampler	0.546	0.499	0.515
	LR + TF-IDF + POS + analyzer = 'char' + ngrams	SMOTE	0.539	0.562	0.542
	MLP + TF-IDF + POS + analyzer = 'char' + ngrams	SMOTE	0.535	0.518	0.515
	LR + TF-IDF + POS + verbs + analyzer = 'word'	R. Over Sampler	**0.565**	0.537	0.544
	LR + TF-IDF + POS + verbs + analyzer = 'word'	SMOTETomek	0.554	0.543	0.542
	RF + Scaler + POS + ngram + LDA Topic modeling	SMOTETomek	0.473	0.466	0.464
	LR + TF-IDF + POS + analyzer = 'word' + ngrams	R. Over Sampler	0.549	0.529	0.532
	SVM + TF-IDF + POS + analyzer = 'char' + ngrams	R. Over Sampler	0.523	**0.695**	**0.586**
	MLP + TF-IDF + POS + verbs + analyzer = 'word'	TomekLinks	0.538	0.593	0.555
Claim-Supporting	LR + Countvectorizer + POS + verbs	SMOTE	0.511	0.533	0.513
	LR + TF-IDF + POS + verbs + analyzer = 'word'	SMOTE	**0.557**	0.520	0.529
	LR + TF-IDF + POS + verbs + analyzer = 'word'	R. Under Sampler	**0.552**	0.450	0.491
	LR + TF-IDF + POS + verbs + analyzer = 'word'	SMOTE	0.541	0.514	0.522
	MLP + TF-IDF + POS + analyzer = 'word' + ngrams	R. Under Sampler	0.545	0.432	0.478
	LR + TF-IDF + POS+ verbs + analyzer = 'word'	TomekLinks	0.540	**0.595**	**0.555**
	LR + TF-IDF + POS + analyzer = 'char' + ngrams	R. Under Sampler	0.506	0.393	0.439
	LR + TF-IDF + POS + analyzer = 'char' + ngrams	SMOTE	0.528	0.455	0.483
	MLP + TF-IDF + POS + analyzer = 'char' + ngrams	SMOTE	0.516	0.510	0.508
	LR + TF-IDF + POS + verbs + analyzer = 'word'	R. Over Sampler	**0.561**	0.536	**0.540**
	LR + TF-IDF + POS + verbs + analyzer = 'word'	SMOTETomek	**0.559**	0.522	0.531
	RF + Scaler + POS + ngram + LDA Topic modeling	SMOTETomek	0.408	0.374	0.385
	LR + TF-IDF + POS + analyzer = 'word' + ngrams	R. Over Sampler	**0.551**	0.546	**0.541**
	SVM + TF-IDF + POS + analyzer = 'char' + ngrams	R. Over Sampler	0.530	0.522	0.515
	MLP + TF-IDF + POS + verbs + analyzer = 'word'	TomekLinks	0.533	**0.590**	**0.547**

by gradually valuating the relative strength of their supporting and attacking arguments to reach rationale-based requirements decisions for the next release of the market-driven software application. For this purpose, we need to identify the sentiments (positive, negative, or neutral) associated with those crowd-user comments in the corpus, which are identified as claim rationale elements.

5.2.1 | Setup for Sentiment Analysis

For this ML experiment, we employ the same classifiers and experimental setup used in the previous experiment, as described in section 5.1.1. In this experiment, the claims rationale element is further classified into supporting-claims, attacking-claims, or neutral-claims to identify conflict-free new features or issues and their associated strength values. The ML experiment is conducted in the same Python environment.

5.2.2 | Results for Sentiment analysis classification

The average 10-fold cross-validation results for different ML algorithms to identify the sentiments of user comments in the Reddit forum by classifying the captured claims rationale element into claim-attacking, claim-supporting, and claim-neutral rationale elements are shown in Table 7 . The classification results are presented for each sentiment type to combine textual features with the corresponding ML algorithm that gives the best results. The experimental results show that the ML classifiers perform better with the data re-balancing approaches, i-e, under-sampling and over-sampling approaches. However, in some cases, under-sampling techniques perform relatively better than the over-sampling approaches, i-e, "TomekLinks" gives the best recall and F1 values with "LR and MLP" algorithms for classifying claim-supporting rationale elements. Furthermore, the hybrid re-balancing technique "SMOTETomek" performed better with the MLP and LR classifiers when classifying the "claims" rationale element into claim-attacking, claim-supporting, and claim-neutral rationale elements. In contrast, we did not get better results with the "SMOTEENN" hybrid re-balancing technique. Hence, we didn't report the results.

In a nutshell, the **LR** and **MLP** classifiers outer-performs the other ML classifiers by achieving better recall, precision, and f-measure in identifying the sentiments of crowd-user comments in the Reddit forum. Both **MLP** and **LR** classifiers gives the best average F-measure scores for classifying claim-attacking rationale element, i-e, **0.473** and **0.464** respectively. The F1 values for the claim-attacking rationale element are not much high. One possible reason might be that the attacking-claims are only 543, compared to neutral-claims and supporting-claims, which are 777 and 742, respectively. For claim-neutral rationale element, the **SVM** classifier gives the highest F-measure value, i-e **0.586**, followed by the **LR** and **MLP** algorithms that gives **0.561** and **0.555**. Finally, similar to the claim-attacking rationale element, both **LR** and **MLP** classifiers gives the best average f-measure scores for classifying claim-supporting rationale element while using the under-sampling re-balancing approach (TomekLink), i-e, **0.555**, **0.547** respectively. Furthermore, we comparatively get better results when employing LDA topic modeling as a textual feature with the different ML to identify the sentiments associated with the crowd-user comments in the Reddit forum by classifying the claims rationale elements into claim-supporting, claim-attacking, and claim-neutral rationale elements, as compared to the classification of user comments to claims, new feature, and issue rationale elements. We showed a classification result for each sentiment element by employing an ML algorithm with LDA topic modeling textual feature in Table 7 , whereas the results obtained are comparatively degraded to the results obtained with other ML algorithms with distant textual features. Amongst the different ML algorithms, the **RF** classifier gives the best results with the LDA topic modeling textual feature, as shown in Table 7 . Based on the experimental results as shown in Table 6 , we can choose either **LR** or **MLP** classifier as a best algorithm to classify crowd-user comments in the Reddit forum into claim-attacking, claim-supporting, or claim-neutral rationale elements. The best textual features with the corresponding ML algorithm that results in the best classification results are depicted in Table 6 .

6 | AUTOMATED VALUATION-BASE REQUIREMENTS ARGUMENTATION AND DECISION-MAKING

Although the crowd-user comments in the Reddit forums may represent a small subset of end-users that would frequently contribute to the development and improvement of the market-driven software, certain crowd-users contribute rationale and requirements-related information in the Reddit forum more often. It has proven that analysis of crowd-users feedbacks on the social media platforms can help software and requirements analysts to capture the conflicting viewpoints and make informed

Algorithm 1 An Algorithm for Automated CrowdRE-VArg

Input: a Tree of user comments with n nodes
Output: Strength value for each identified new feature, issue, or main discussion argument

```
function COMMENTS_AGGREGATION(child_tree)
    child_node_values = []
    if child_tree == [] then
        total_aggregation_child = 0
        return total_aggregation_child
    else
        for child in child_tree do
            Father_tree= Get the parent-tree of the current child
            for node in range(0, len(Father_tree)) do
                if node == (len(Father_tree)-1) then
                    child_node_values.append(Father_tree[node])
                end if
            end for
            aggregation_child = 1
            for value in child_node_values do
                aggregation_child = aggregation_child * (1 - value)
            end for
            total_aggregation_child = 1 - aggregation_child
            return total_aggregation_child
        end for
    end if
end function
function COMMENTS_COMBINATION(Vs, Va)
    BS= 0.5
    if Va == Vs then
        child_tree_N= get supporting and attacking nodes of "N"
        Update_child_tree= ASSIGN_VALUATION(child_tree_N,BS)
    end if
    if Va > Vs then
        sub_combination = Vs -Va
        if sub_combination < 0 then
            sub_combination = sub_combination * (-1)
        end if
        combination = BS - (BS * sub_combination)
        child_tree_N= get supporting and attacking nodes of "N"
        Update_child_tree= ASSIGN_VALUATION(child_tree_N,BS)
    end if
    if Va < Vs then
        sub_combination = Vs -Va
        if sub_combination < 0 then
            sub_combination = sub_combination * (-1)
        end if
        combination = BS + ((1 - BS) * sub_combination)
        child_tree_N= get supporting and attacking nodes of "N"
        Update_child_tree= ASSIGN_VALUATION(child_tree_N,BS)
    end if
end function
function ASSIGN_VALUATION(child_tree, valuation)
    for child in range(0, len(child_tree)) do
        if valuation != None then
            if child == (len(child_tree)-1) then
                child_tree[child]= valuation
                return child_tree
            end if
        else
            if child == (len(child_tree)-1) then
                valuation = child_tree[child]
                return valuation
            end if
        end if
    end for
end function
father_children = Get Level Order Traversal of Tree O
for each node N in O do
    S_Set[node] = {N's child nodes who support N}
    N_Set[node] = {N's child nodes who attack N}
    S_N_Set[node] = {N's child nodes of parent node N}
```

```
end for
for each node N in S_N_Set do
    append the Base Score (BS) to each node of S_N_Set
end for
T_reverse = reverse the tree of user comments (O)
for each node N in T_reverse do
    // C is a node classifier
    if C(N) == 'issue' or C(N) == 'new feature' then
        child_tree = Get child tree of Node N
        len_child_tree= len(child_tree)
        if len_child_tree <= 1 then
            print the total strength of the new feature of issue as en_child_tree[0]
        else
            get_S_childs = get supporting nodes of "N"
            get_N_childs = get attacking nodes of "N"
            total_aggregation_S= COMMENTS_AGGREGATION(get_S_childs)
            total_aggregation_A= COMMENTS_AGGREGATION(get_N_childs)
            Vs = total_aggregation_S, Va= total_aggregation_N
            COMMENTS_COMBINATION(Vs, Va)
            child_tree = Get child tree of Node N
            valuation = None
            child_tree= ASSIGN_VALUATION(child_tree,valuation)
            print the total strength of the new feature of issue
        end if
    else
        if C(N) == 'claim' then
            f = Get value N.parent [N]
            if  N in father_children then
                if R(f,N) == 'attack' or R(f,N) == 'support' or N == root node then
                    get_S_childs = get supporting nodes of "N"
                    get_N_childs = get attacking nodes of "N"
                    total_aggregation_S= COMMENTS_AGGREGATION(get_S_childs)
                    total_aggregation_A= COMMENTS_AGGREGATION(get_N_childs)
                end if
                Vs = total_aggregation_S, Va= total_aggregation_N
                COMMENTS_COMBINATION(Vs, Va)
                if N == root node then
                    valuation = None
                    child_tree= ASSIGN_VALUATION(child_tree,valuation)
                    print the total strength of the root node
                end if
            end if
        end if
    end if
end for
```

requirements decision-making[27]. However, manually processing the crowd-users feedback in the Reddit forum to reach a certain consensus is a tedious and time-consuming task due to the large amount of user feedback. For this purpose, we proposed an automated CrowdRE-VArg approach, which helps to negotiate the conflicts over the identified new features or issues between the different crowd-users on the run in the ongoing discussion by gradually valuating their supporting attacking arguments. To this end, in the below sub-sections, we introduce an automated CrowdRE-VArg approach and a case study of actual crowd-users discussion in the Reddit forum used to elaborate on the proposed automated approach by finding a settlement that would satisfy the involved crowd-users. The proposed CrowdRE-VArg approach requires the user conversation in the Reddit forum to be converted into an argumentation tree to process each node and make certain requirements-related decisions underneath the rationale. For this purpose, we choose an algorithm from the past work[27], which takes n-number of end-users feedback from the user conversation in the Reddit forum as input and after processing outputs as an argumentation tree.

6.1 | Algorithm to automate CrowdRE-VArg framework

The algorithm. 1 automatically identifies the strength of the captured new features or issues by gradually valuating their supporting and attacking arguments. The algorithm implements the proposed CrowdRE-VArg approach and semantics inspired by GVBA, QuAD, ESAAF, and DF-QuAD. It gives a prioritized list of new features or issues by identifying the strength value of each rationale element captured against the main discussion topic. Also, it captures the strength value of the main discussion topic by valuating its supporting and attacking arguments helping software and requirements engineers to understand the crowd-users opinions about the ongoing discussion topic. The algorithm takes a "crowd-user comments tree" of infinite depth and breadth as an input while returning the strength value with each identified new feature, issue, or the main discussion topic. For this purpose, an n-level-tree (both depth and breadth) is constructed automatically from the end-users feedback based on the "Comment_Id", and "Parent_Id", which will be saved as a Python Dictionary, where each node and sub-node will be represented as a "Key", and the children of the corresponding node or sub-node will be represented as "Values". The automated tree of the crowd-users comments from in the Reddit forum is constructed with an algorithm proposed in our previous work[27]. Three separate python dictionaries are created that contain supporting arguments, attacking arguments, and both supporting and attacking arguments, respectively, for each node in the argumentation tree, if they have any. According to the proposed CrowdRE-VArg, a BS of "0.5" is assigned with each node in the argumentation tree, which is considered pivotal in requirements decision-making, particularly in requirements prioritization and negotiations. As discussed earlier, the CrowdRE-VArg framework works on the principle of a bottom-up approach; therefore, the argumentation tree of the crowd-user comments is reversed. The Algorithm 1, is comprised of three main functions, i-e, "COMMENTS_AGGREGATION", "COMMENTS_COMBINATION", and "ASSIGN_VALUATION". The "COMMENTS_AGGREGATION" function, implements the arguments aggregation function, as described in equation 1. To identify the strength of a new feature, issue, or the main discussion topic, we separately call the "COMMENTS_AGGREGATION" function to identify the aggregated value of their supporting and attacking arguments for each identified new feature, issue, or the main discussion topic, as explained in equation 1. After identifying the aggregated supporting and attacking values for each new feature, issue, or the main discussion topic, we call the "COMMENTS_COMBINATION" function to identify their dialectical strength based on their corresponding aggregated supporting and attacking strength values. The "COMMENTS_COMBINATION" function implements the equation 2. While the function "ASSIGN_VALUATION" assigns the actual strength value to the captured new feature, issue, or the main discussion topic identified through the aggregation and combination functions. Finally, the algorithm returns the new features, issues, or the main discussion topic with the accumulated strength values by gradually valuating their supporting and attacking arguments in the argumentation tree. As a result, one can get the prioritized list of new features, issues, or the main discussion topic with their aggregated strength values, based on which software developers and requirements engineers can take certain requirements-related decisions by considering the crowd-user opinions.

6.2 | Elaborating the proposed CrowdRE-VArg algorithm with a running example

To elaborate on the automated CrowdRE-VArg approach, we take the first discussion topics "Google Maps is testing a combined commute to replace driving and transit" about the Google map mobile application from the Reddit forum as an example case (case study), shown in the Table. 2 . It works as a proof-of-concept by highlighting the advantages and feasibilities of the proposed automated CrowdRE-VArg approach for the software and requirements engineers. Still, the proposed approach ML experimental results need improvements as misclassification can affect the required results because of the bottom-up approach. Also, we selected a constant $\mathcal{BS}$ for the CrowdRE-VArg approach, and some arguments in the Reddit forum are more powerful and important and may be assigned a higher $\mathcal{BS}$ and vice-versa. For this, we can utilize the ASPIC+ argumentation framework[51] to identify the end-user comment strength and combine the value with the $\mathcal{BS}$ to improve the validity and effectiveness of the

CrowdRE-VArg approach. However, when experimenting with the main discussion topic annotated manually, the CrowdRE-VArg approach identified the strength values of the conflict-free new features and issues by gradually valuating their pro and con arguments. The automated approach significantly reduces the efforts required to identify the new features and issues with their aggregated scores compared to the manual approach. The approach helps in improving the quality of the software applications by incorporating the in-time feedback from the end-user in the Reddit forum.

When running the automated CrowdRE-VArg approach, it identifies 23 new features, amongst which 17 receive a default strength value of "0.5" (in compliance with Rule No. 3 of CrowdRE-VArg) because other crowd-users did not register supporting or attacking arguments against them. For example, "I just want to see the street names", "I wish they had something for truckers, like a truck route only section", etc. The new feature "I wish Google Maps was slightly less US-centric in its feature set. Much of the world cycle commutes. Even some Americans" earn the highest strength value "0.8125" by valuating their corresponding supporting and attacking arguments. The detailed demonstration of how it is calculated is depicted in section 4.2. Thus, the proposed CrowdRE-VArg approach makes it simple and easy for the requirements and software engineers to identify conflict-free requirements-related pivotal information and hide the complex theory behind it. Similarly, two new features get the second-highest strength values of "0.75", against which other end-users registered a single supporting argument. Similarly, the new feature "can't they just rename the tab dynamically based on whether you selected drive or public transport in the 'how to get to work?' question they asked you? this is how Google Assistant handles things", earns a strength value of "0.252" by gradually valuating a series of supporting and attacking arguments. Its strength value is the second least among the identified new features in the discussion topic. Its pictorial representation with its supporting and attacking arguments is shown in Figure. 3 . Finally, a new feature or alternative feature, "Citymapper does that, just set your commute start point to your park-and-ride station, and it will tell you when you have to get to that station every morning to get to work in time" gets the least strength value of "0.25", as another crowd-user registered an attacking argument against it in the Reddit forum.

Similarly, the automated CrowdRE-VArg approach identifies four issues in the discussion topic, where each captured issue receives a default strength value of "0.5" (in compliance with Rule No. 3) because other crowd-users didn't register supporting or attacking arguments against them. For example, "Why does Google Maps have tabs in the first place?", "huh? it has a cycling option", etc. Also, the automated CrowdRE-VArg approach identifies the strength value of the main discussion topic by valuating its supporting and attacking arguments. For the first discussion topic, i-e, "Google Maps is testing a combined commute tab to replace Driving and Transit", the automated CrowdRE-VArg approach identifies the strength value of "0.502" by gradually valuating the four supporting and three attacking arguments, against which other end-users gave supporting or attacking arguments. For example, its supporting arguments are: "I like it", "Good, keep it simple", etc. While its attacking arguments are: "Maps is getting way too bloated and uses way too much rich-media. The explore tab could be its own app" that gets a single supporting argument "Way too bloated, I only use it to locate a place but rarely to do anything about the location, got better apps for that.", "i don't know how practical it would be, especially for people that don't drive or the others that don't take public transits" against which other end-users gave eleven attacking and one supporting argument, etc. This useful information for the software developers and requirements engineers are automatically identified when taking certain requirements-related decisions in selecting new features or improving existing features by considering the hot issues identified for the next release of the software application under discussion. Additionally, the automated approach document the rationale for the various requirements decision making the process transparent and user-centric by considering the end-user opinions associated with the captured new features or issues in the Reddit forum. However, employing the automated CrowdRE-VArg approach on larger discussion topics in the Reddit forum still requires handsome improvements in the results of the ML classifications to minimize the effect of misclassifications arrived due to the scarcity of the input data. Therefore, the automated CrowdRE-VArg approach might be tested and validated with larger similar end-user data from the Reddit forum. Then, state-of-the-art deep learning (DL) algorithms might be used to get better classification results and improve the overall performance of the CrowdRE-VArg approach.

6.3 | Analysis of the proposed CrowdRE-VArg Algorithm

The main building block of the proposed approach lies in the data set, which it uses as an input that comes through the annotation process. As explained in section 3.5, it performs better when the standard procedures are followed for the data annotation. However, it is considered one of the possible limitations of the CrowdRE-VArg approach that it operates on a specially curated and annotated data set, i-e, the automated algorithm uses "Comment_Id", and "Parent_Id" to recognize the possible relation between the two follow-up crowd-user comments, which can be overcome and improved by introducing argumentation mining (AM)[73] and DL algorithms, such as LSTM to detect the possible relationship between the two end-users comments automatically[19]. In section 6.2, we highlighted the advantages and feasibilities of the proposed automated CrowdRE-VArg approach for the software and requirements engineers. However, to evaluate the automated CrowdRE-VArg approach, we have taken certain assumptions that might affect the operation and efficiency of the proposed automated CrowdRE-VArg approach and its algorithm. Additionally, it could help to identify possible anomalies in the current state-of-the-art user forums and how we can fix and improve them to support argumentation-based requirements acquisition, which in turn help requirements analysts in making certain decisions for the future release of the software application under consideration.

While manually analyzing the user comments on the Reddit forum, it is observed that crowd-users submit useful information about the software application under discussion, i-e, suggest new features, ask for quality improvement, or report issues encountered by them using the current software application. Such information is pivotal for requirements engineers in making requirements-related decision-making, but this remains often hidden from the crowd-users and software developers due to the many crowd-user comments submitted in the Reddit forum. Also, the crowd-user comments that gain high votes are shown first in the discussion topic and vice-versa. In the process, the useful requirements-related information might remain hidden due to the possible skewed voting or any other biasness, i-e, the social influence of certain crowd-users, in the Reddit forum. For this purpose, the proposed CrowdRE-VArg approach provides an opportunity to valuate each user comment based on their supporting and attacking arguments in the Reddit forum with the $\mathcal{BS}$ to identify the strength value of each user comment. It would help to bring upfront the useful requirements-related information, i-e, user comments with lower voting, by modifying the current interface of the Reddit forum. For this purpose, employing simple yet more instinctive statistics on the alternatives, new features, or issues captured in the ongoing user conversations in the Reddit forum underneath rationale. One can employ the contingency tables or mosaic plots to provide an overview that highlights the rationale and requirements-related elements in the end-user conversation and contains the argumentation and reasoning to have a clearer image for the incoming crowd-users about the ongoing discussion topics in the Reddit forums.

We assume that the crowd-users in the Reddit forum did not get equal opportunities because the user conversations in the forum are offline, i-e, there is no restriction on the end-users to conclude the ongoing discussion, might be a power disruption or connectivity issue when participating in the end-user discussions. Therefore, the time frame might affect the strength values of the captured new features, issues, or the main discussion topic identified through the CrowdRE-VArg approach in the Reddit forum. For example, introducing a new supporting or attacking argument will change the strength value of the captured new features, issues, or the main discussion topic, which remained unposted due to the power failure or connection loss. However, we can control the validity of the automated CrowdRE-VArg approach by introducing a design change in the interface of the current Reddit forum, i-e, adding a timestamp (both to and from) that will help control the strength values of the identified requirements-related elements by controlling the timestamp. We also assume that the software rationale might help end-users understand the complexity of the software application under discussion and comprehend the tradeoffs that users might not have to think about, which might enhance the software application ratings. Currently, the proposed automated CrowdRE-VArg approach does not visualize the supporting and attacking arguments highlighting conflicting user opinions on the main discussion topics, new features or issues in the Reddit forum. Also, the algorithm can't handle the different ways to motivate crowd-users to participate in contributing feedback in the Reddit forum. However, we can control this validity by restructuring the ongoing discussion in the Reddit forum, which might improve the crowd-users engagements in providing valuable feedback and visualizing the pro and con argument for the identified requirements-related elements in the Reddit forum. For example, we can display the profiles of the key stakeholders in the Reddit forum who are frequently contributing rationale and requirements-related information, i-e, new features, issues, or design alternatives. Also, we can boost the end-user motivations in providing quality feedback by giving various incentives, i-e, social recognition, cash, or vouchers.

7 | DISCUSSION

This section discusses the importance of argumentation theory and rationale management for the Crowd-based requirements engineering to achieve informed requirements decisions. Below we discuss and summarizes the research finding in detail:

7.1 | CrowdRE-VArg for requirements elicitation

Our previous research findings reveal that almost 71% of the crowd-user comments in the Reddit forum contain rationale and requirements-related information[28]. In comparison, a key challenge in market-driven, distributed software development projects is to engage different stakeholders in various RE activities with minimal effort. For instance, capturing, identifying, and avoiding the possible conflicts between the culturally and physically distributed stakeholders in finalizing certain features for the next release of the software application. It is a key issue in market-driven collaborative RE, where software application requirements for the next release or important issues in the current version are elicited in a distributed process from geographically distributed stakeholders. Different social media platforms enable such relevant information for the requirements engineers, i-e, user forums, app stores, etc.[18]. In these social media communities, mostly free natural language text engages different stakeholders in discussions or shares their experience or knowledge with the software application.

Recently, its becomes of pivotal importance for CrowdRE to develop capable techniques that 1) automatically identify and capture relevant requirements and rationale information from the various social media platforms and 2) automatically filter out conflicting viewpoints in the user conversations on these platforms to have informed requirements decisions-making. Keeping this in view, we proposed the CrowdRE-VArg approach that automatically analyses the crowd-users conversations in the Reddit forum and identifies the strength values of the requirements-related elements, i-e, new features or issues, by gradually valuating their supporting or opposing arguments in the Reddit forum. Such requirements-related information is useful for requirements engineers utilized for a future release of software application[3,65]. Also, the issues reported by the crowd-user are often helpful for requirements engineers to understand the level of difficulty faced by the end-users while using a software application[65,3,25].

7.2 | Insights for Requirements and software Engineers with CrowdRE-VArg

While the recent research focuses on mining rationale information from text documents and social media platforms, the CrowdRE-VArg approach takes a different perspective to support software and requirements engineers in requirements decision-making by identifying the strength values of conflict-free new features or issues and documents the rationale behind. The research findings of the proposed CrowdRE-VArg approach show that the Reddit forum can be treated as a rich source for requirements elicitation, prioritization, and software rationale mining. It supports the requirements elicitation and prioritization activities by capturing a prioritized list of new features, issues, or design alternatives with respective strength values. Also, it helps requirements analysts negotiate the conflicts over the new features or issues between the geographically distributed crowd-users on the run by finding a settlement that satisfies the involved users in the Reddit forum. Capturing user rationale in the Reddit forums also supports software developers and requirements analysts in filtering user feedback and making better requirements decisions. For example, during requirements analysis, capturing winning requirements-related elements by identifying their strength values and improving documentation referred to when the development team changes. Software user rationale is beneficial in marking and filtering potentially high or low informative user feedback on social media platforms. Those user reviews that do not contain rationale information can be filtered out as less informative reviews. They can be further explored to understand why a user switches to a software application by exploring the various issues faced while using the software application.

Another perspective of the CrowdRE-VArg approach is that software developers can benefit from the user rationale when prioritizing and negotiating new features or issues for the latest release or improving the existing version. For example, a new software feature or an issue captured might be given higher priority when many users record supporting arguments in response or they are mentioned too frequently in the user comments. Comparative studies and analysis between the two software applications might help software practitioners, and developers understand how better their application can perform compared to the existing most mentioned alternatives software applications. Additionally, software developers can summarize or group similar requirements into the same clusters by exploring different argumentation frameworks, i-e, AAF[21], Weighted abstract argumentation[74], and BAF[34], to identify the user comments or requirements in a textual document or social media platform that need to be included in the textual summary or a cluster. It will work on the principle of abstract argumentation, where attack relation indicates that two user comments cannot be both selected for the textual summary or inside a cluster. Similarly, support relations indicate that two user comments or requirements should be both selected for the textual summary or the same cluster[75]. We can generate multiple graphs or mockups for the software developers using the proposed CrowdRE-VArg approach by analyzing the user feedback to support requirements and design decision-making. For this purpose, we can show the supporting and attacking arguments against each identified new feature or issue with the count in the Reddit forum, and based on these arguments; we can make a graph with two curves representing positive and negative comments against the identified rationale element with time. The software developers can also use this trend to prioritize the identified new features and issues. For example, suppose the number of positive user comments extends more than negative comments for the captured new feature or issue rationale element. It can be assigned a larger value when prioritizing them for the next release of the market-driven software application. Moreover, we can apply the proposed CrowdRE-VArg approach to a transcribed focus group conversation to capture the conflicting viewpoints and identify the alternative solutions if the nature of the discussion is textual.

7.3 | Cognitive support for Crowd-users

The large volume of user feedback from different social media platforms, such as user forums (particularly the Reddit forum), Amazon's online store, uservoice.com, google, and app stores, provide a pivotal source for the software companies to examine the multiple viewpoints and the broad range of software rationale in the end-users discussion. On the other side, many such reviews challenge the end-users to quickly and efficiently survey the growing amount of information and provide feedback to get involved in the ongoing discussion. Incorporating user rationale in the Reddit forum can support user involvement by restructuring the ongoing user discussion. Likes, a simple statistic over the user reviews in the Reddit forum, can provide an overview of the proposed new features, issues identified, and their corresponding strength values based on supporting and attacking arguments. These simple statistics will improve the user involvement in the ongoing discussion and encourage other users to participate in the ongoing conversation, improving the overall rating of the software application.

8 | THREATS TO VALIDITY

This section discusses the internal and external validities of the proposed CrowdRE-VArg approach. The details are given below:

Internal threats: For the proposed CrowdRE-VArg approach, the authors of the research article who annotated and labeled the crowd-user comments in the data set were also involved in the experiment design and evaluation. The human coders involved in the annotation process are subject to errors, just like any other manual annotation study. Also, there is a probability that the annotation experts have subconsciously attempted a second guess leading to disagreements. However, **to mitigate the risk**, the authors carried out the crowd-user comments annotation in a more professional, systematic, and iterative manner. Two annotators with several years of research and development experience individually annotated end-user reviews. Furthermore, a coding guideline is developed by considering the best practices to verify and validate the annotation process.

Another threat to the internal validity is the taxonomy we selected for the proposed CrowdRE-VArg approach to classify the end-user comments in the Reddit forum. This research study focuses on four types of information considered relevant to the software developers and proposed CrowdRE-VArg approach. **To alleviate this threat**, we selected an initial set of categories by critically analyzing a random sample of end-user reviews in the Reddit forum and observing the previously published related work. We assume these four categories are considered as the most relevant to the requirements engineering research community. Although, we believe that one important category, i.e., non-functional requirements, has been missed due to its most minor occurrence in the data set collected from the Reddit forum. Also, we might have missed some promising ML techniques or textual features that could enhance the accuracy of the ML algorithms used in the experiment. We, therefore, did not claim completeness in the classifier's accuracy, as this research study was exploratory. However, we introduced and incorporated techniques and textual features that have been proven successful in the past research work.

External Threats: A possible external threat is that the proposed CrowdRE-VArg approach is not aimed to be generalized nor representatives of other popular social media platforms, such as app store reviews, amazon user reviews, and Twitter data. We believe that our results are symbolic and are a promising first step into the previously undiscovered Reddit forums to identify conflicting viewpoints and arguments strength using argumentation theory. However, we can get different results when applying the proposed approach to other social media-generated data sets because they have different structures and meta-data than the Reddit forum used in this research study. **To overcome this threat**, we are interested in employing AM[73], DL, and textual entailment[76] to generalize the proposed CrowdRE-VArg approach and test it on the data set curated from other social media platforms, i-e, Twitter and app stores.

Another possible external threat of the CrowdRE-VArg approach is, for now, we treat each end-user comment in the Reddit forum as of equal strength when identifying the strength values of the identified new features, issues, or the main discussion topic. Still, some crowd-users comments might be stronger than other user comments on the online discussion topics. **To resolve this threat**, we plan to introduce the ASPIC+ argumentation framework[51] to understand the internal structure and strength of user comments and improve the strength values of the identified requirements elements. Also, for the CrowdRE-VArg approach, we have selected a constant BS, while some arguments in the Reddit forum are more powerful that may be assigned a higher

$\mathcal{BS}$ and vice-versa. **To address this threat**, we can identify the strength of each user argument in the Reddit forum by using the ASPIC+ framework[51] and combine the value with the $\mathcal{BS}$ to resolve this threat. Furthermore, we believe that the user comments data set employed in the proposed CrowdRE-VArg approach is rather small, which is considered as a possible threat. To resolve this threat, we needs to collect user comments on a larger scale by grouping the relevant topics in the Reddit forum about the Google Maps mobile application to expedite the state-of-the-art DL algorithms together with the proposed CrowdRE-VArg approach to improve the classification results and overall performance of the CrowdRE-VArg approach.

9 | CONCLUSION AND FUTURE WORK

Argumentation-based rationale and management play a pivotal role in software and requirement engineering. The rationale knowledge is considered a primary element of the requirements and design knowledge because many important decisions are taken during these phases, which must be efficiently, and effectively captured, managed, and documented. Various rationale representation models have been proposed to identify and capture rationale elements in the textual documents with varying degrees of success in practice. However, because of the complex rationale structure and less tool support, their use in practice has been limited and hindered. Recently, software engineering researchers have focused on automated mining of rationale information from software artifacts and social media platforms to capture rationale knowledge from textual documents. Such automated approaches help to significantly minimize and reduce the cost associated with the rationale identification, capturing, retrieving, and managing, hence improving the stakeholder's involvement in the application development[77].

The finding of the proposed CrowdRE-VArg approach shows that the Reddit forum can be used as a rich source for requirements opinions elicitation and rationale mining. It supports requirements elicitation activity by identifying new features or issues. For this purpose, firstly, we carried out a detailed analysis of the end-user comments on Reddit.com to identify and capture the different rationale elements that help in the requirements decision-making for the market-based software evolution. The key rationale and requirements elements of interest include design alternatives, new features, pro or con arguments, and issues or challenges. Secondly, we collected 3123 crowd-user comments from the seven online discussion topics about the Google Maps mobile application in the Reddit forum. The first discussion topic shown in Table. 2 is used as a proof-of-concept (case study) for the proposed CrowdRE-VArg approach. In contrast, the remaining six discussion topics (3046 user comments) are used as an experimental data set to conduct fine-grained machine learning and sentiment analysis experiments. Next, we proposed the CrowdRE-VArg approach, inspired by GVBA, QuAD, ESAAF, and DF-QuAD, to identify conflict-free design alternatives, new features, and issues to be considered by the software engineers for coming release based on their pro and cons arguments. The crowd-user feedback in the Reddit forum is first converted into an argumentation tree. Then, each end-user comment in the argumentation tree is processed to identify each new feature's strength value, issue, or main discussion topic by valuating their pro and con arguments. Besides, to automate the CrowdRE-VArg, an algorithm is developed that takes n-number of user comments and returns the strength values for each requirements-related element and the main discussion topic. Finally, to scale the proposed CrowdRE-VArg approach, we employ different ML algorithms that classify user comments into new features, issues, and claim rationale elements. Then employee sentiments analysis approach over the user comments identified as claim rationale element to identify end-user opinions (supporting, attacking, or neutral) about the new feature, issue, or the main discussion topic.

AM is a new research direction for natural language text processing. It automatically recognizes the argumentative structure in a text document, i-e, the conclusion or assumption, premises, and the complete arguments; even it helps to identify the relationships between the arguments[73]. In the future, we can overcome the manual annotation of end-user comments in the Reddit forum by introducing AM and DL algorithms, such as LSTM, to automatically detect the possible relationship between the two end-users comments[19]. In the future, we can identify the strength of each user argument in the Reddit forum by using ASPIC+ argumentation framework[51] and then combine the argument strength value with the base score to improve the results of the proposed CrowdRE-VArg approach. Furthermore, We can optimize the CrowdRE-VArg approach by improving the performance of the machine learning algorithms in identifying the rationale elements in the user conversation by exploring computational linguistics. For this, we need to intertwine argumentation reasoning with supervised machine learning algorithms[56]. The proposed modified CrowdRE-VArg approach can be utilized for the software acquisition problem to select the best-suited software vendors based on evaluating their pro and con arguments collected from app stores and Twitter[78]. It is concluded that AM is still a young research field concerning opinion mining, which has been heavily applied by the software and requirements engineering researchers to identify and capture meaningful and actionable information types from the crowd-users comments on social media. Although some research work is present that leverage certain contributions from AM research to mine rationale information from the software artifacts, there is still a gap in the application of AM approaches for rationale and requirements mining from the user review in the social media platforms.

Data Availability: Enquiries about data availability should be directed to the authors.

References

1. Panichella S, Di Sorbo A, Guzman E, Visaggio CA, Canfora G, Gall HC. How can i improve my app? Classifying user reviews for software maintenance and evolution. *2015 IEEE 31st International Conference on Software Maintenance and Evolution, ICSME 2015 - Proceedings* 2015(1): 281–290. doi: 10.1109/ICSM.2015.7332474

2. Ali Khan J, Liu L, Jia Y, Wen L. Linguistic Analysis of Crowd Requirements: An Experimental Study. *Proceedings - 2018 7th Workshop on Empirical Requirements Engineering, EmpiRE 2018* 2018: 24–31.

3. Khan JA, Xie Y, Liu L, Wen L. Analysis of requirements-related arguments in user forums. *Proceedings of the IEEE International Conference on Requirements Engineering* 2019; 2019-September: 63–74.

4. Liu C, Yang W, Li Z, Yu Y. Recommending software features to designers: From the perspective of users. *Software: Practice and Experience* 2020; 50(9): 1778–1792.

5. Liu H, Zhang M, Liu L, Liu Z. A method to acquire cross-domain requirements based on Syntax Direct Technique. *Software: Practice and Experience* June 2021; 51(8).

6. Maalej W, Nayebi M, Johann T, Ruhe G. Toward data-driven requirements engineering. *IEEE Software* 2016; 33(1): 48–54.

7. Bakiu E, Guzman E. Which feature is unusable? Detecting usability and user experience issues from user reviews. *Proceedings - 2017 IEEE 25th International Requirements Engineering Conference Workshops, REW 2017* 2017: 182–187.

8. Wang W, Arya D, Novielli N, Cheng J, Guo JL. ArguLens: Anatomy of Community Opinions On Usability Issues Using Argumentation Models. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* 2020: 1–14.

9. Kurtanović Z, Maalej W. On user rationale in software engineering. *Requirements Engineering* 2018; 23(3): 1–23.

10. Alkadhi R, Nonnenmacher M, Guzman E, Bruegge B. How do developers discuss rationale?. *2018 IEEE 25th International Conference on Software Analysis, Evolution and Reengineering (SANER)* 2018: 357–369.

11. Rogers B, Justice C, Mathur T, Burge JE. Generalizability of document features for identifying rationale. *Design Computing and Cognition'16* 2017: 633–651.

12. Dutoit AH, Paech B. Rationale management in software engineering. *Handbook of Software Engineering and Knowledge Engineering: Volume I: Fundamentals* 2001: 787–815.

13. Jarczyk AP, Löffler P, Shipman F. Design rationale for software engineering: a survey. *Proceedings of the Hawaii International Conference on System Sciences* 1992; 25: 577–577.

14. Dutoit AH, Paech B. Rationale-based use case specification. *Requirements engineering* 2002; 7(1): 3–19.

15. ASME . Responses to "Question of the month,December 2004". *Mechanical Engineering Magazine* 2005: 10.

16. Burge JE, Brown DC. Software engineering using rationale. *Journal of Systems and Software* 2008; 81(3): 395–413.

17. Kasauli R, Knauss E, Horkoff J, Liebel G, Oliveira Neto dFG. Requirements engineering challenges and practices in large-scale agile system development. *Journal of Systems and Software* 2021; 172: 110851.

18. Khan JA, Liu L, Wen L, Ali R. Crowd intelligence in requirements engineering: Current status and future directions. *International working conference on requirements engineering: Foundation for software quality* 2019: 245–261.

19. Bosc T, Cabrio E, Villata S. Tweeties Squabbling: Positive and Negative Results in Applying Argument Mining on Social Media.. *COMMA* 2016; 2016: 21–32.

20. Cocarascu O, Toni F. Mining bipolar argumentation frameworks from natural language text. *Argument* 2017; 2(t3): t1.

21. Dung PM. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence* 1995; 77(2): 321–357.

22. Elrakaiby Y, Ferrari A, Spoletini P, Gnesi S, Nuseibeh B. Using argumentation to explain ambiguity in requirements elicitation interviews. *2017 IEEE 25th International Requirements Engineering Conference (RE)* 2017: 51–60.

23. Bagheri E, Ensan F. Consolidating multiple requirement specifications through argumentation. *Proceedings of the 2011 ACM Symposium on Applied Computing* 2011: 659–666.

24. Al-Alshaikh HA, Mirza AA, Alsalamah HA. Extended Rationale-Based Model for Tacit Knowledge Elicitation in Requirements Elicitation Context. *IEEE Access* 2020; 8: 60801–60810.

25. Morales-Ramirez I, Perini A. Argumentation-based discussion for user forum: A Research Preview. *International Working Conference on Requirements Engineering: Foundation for Software Quality* 2014: 232–238.

26. Khan JA, Yasin A, Assam M, et al. Requirements decision-making as a process of Argumentation: A Google Maps Case Study with Goal Model. *International Journal of Innovations in Science & Technology* 2021; 3(4): 15–33.

27. Khan JA, Liu L, Wen L, Ali R. Conceptualising, extracting and analysing requirements arguments in users' forums: The CrowdRE-Arg framework. *Journal of Software: Evolution and Process*: e2309.

28. Khan JA, Liu L, Wen L. Requirements knowledge acquisition from online user forums. *IET Software* 2020; 14(3): 242–253.

29. Rago A, Toni F, Aurisicchio M, Baroni P. Discontinuity-free decision support with quantitative argumentation debates. *Proceeding of International Workshop Temporal Representation and Reasoning* 2016.

30. Cayrol C, Lagasquie-Schiex MC. Gradual valuation for bipolar argumentation frameworks. *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty* 2005: 366–377.

31. Cayrol C, Lagasquie-Schiex MC. Graduality in argumentation. *Journal of Artificial Intelligence Research* 2005; 23: 245–297.

32. Baroni P, Romano M, Toni F, Aurisicchio M, Bertanza G. An argumentation-based approach for automatic evaluation of design debates. *International Workshop on Computational Logic in Multi-Agent Systems* 2013: 340–356.

33. Evripidou V, Toni F. Argumentation and voting for an intelligent user empowering business directory on the web. *International Conference on Web Reasoning and Rule Systems* 2012: 209–212.

34. Cayrol C, Lagasquie-Schiex MC. On the acceptability of arguments in bipolar argumentation frameworks. *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty* 2005: 378–389.

35. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. *Proceedings of the 18th international conference on evaluation and assessment in software engineering* 2014: 1–10.

36. Khan AA, Shameem M, Nadeem M, Akbar MA. Agile trends in Chinese global software development industry: Fuzzy AHP based conceptual mapping. *Applied Soft Computing* 2021; 102: 107090.

37. Haley CB, Moffett JD, Laney R, Nuseibeh B. Arguing security: Validating security requirements using structured argumentation. *In. Proceedings of 3rd Symposium on RE for IS with the 13th International RE Conference* 2005.

38. Franqueira VN, Tun TT, Yu Y, Wieringa R, Nuseibeh B. Risk and argument: A risk-based argumentation method for practical security. *2011 IEEE 19th International Requirements Engineering Conference* 2011: 239–248.

39. Kovacs A, Karakatsanis I, Svetinovic D. Argumentation-based security requirements analysis: Bitmessage case study. *2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom)* 2014: 408–414.

40. Ionita D, Bullee JW, Wieringa RJ. Argumentation-based security requirements elicitation: The next round. *2014 IEEE 1st International Workshop on Evolving Security and Privacy Requirements Engineering (ESPRE)* 2014: 7–12.

41. Tun TT, Bandara AK, Price BA, et al. Privacy arguments: Analysing selective disclosure requirements for mobile applications. *2012 20th IEEE International Requirements Engineering Conference (RE)* 2012: 131–140.

42. Jureta I, Mylopoulos J, Faulkner S. Analysis of multi-party agreement in requirements validation. *2009 17th IEEE International Requirements Engineering Conference* 2009: 57–66.

43. Elrakaiby Y, Ferrari A, Mylopoulos J. Care: A refinement calculus for requirements engineering based on argumentation semantics. *2018 IEEE 26th International Requirements Engineering Conference (RE)* 2018: 364–369.

44. Bhat M, Shumaiev K, Biesdorf A, Hohenstein U, Matthes F. Automatic extraction of design decisions from issue management systems: a machine learning based approach. *European Conference on Software Architecture* 2017: 138–154.

45. Liang Y, Liu Y, Kwong CK, Lee WB. Learning the "Whys": Discovering design rationale using text mining—An algorithm perspective. *Computer-Aided Design* 2012; 44(10): 916–930.

46. López C, Codocedo V, Astudillo H, Cysneiros LM. Bridging the gap between software architecture rationale formalisms and actual architecture documents: An ontology-driven approach. *Science of Computer Programming* 2012; 77(1): 66–80.

47. Nicole Murphy . Reddit's 2019 Year in Review, Highlighting Reddit's top communities, topics and trends of the year. https://www.redditblog.com/2019/12/04/news/reddits-2019-year-in-review, Last accessed on April 2019.

48. Corbin J, Strauss A. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications . 2014.

49. Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit.. *Psychological bulletin* 1968; 70(4): 213.

50. Villata S. Artificial Argumentation for Humans.. *IJCAI* 2018: 5729–5733.

51. Modgil S, Prakken H. The ASPIC+ framework for structured argumentation: a tutorial. *Argument & Computation* 2014; 5(1): 31–62.

52. Leite J, Martins J. Social abstract argumentation. *Twenty-Second International Joint Conference on Artificial Intelligence* 2011.

53. Karlsson J, Wohlin C, Regnell B. An evaluation of methods for prioritizing software requirements. *Information and software technology* 1998; 39(14-15): 939–947.

54. Achimugu P, Selamat A, Ibrahim R, Mahrin MN. A systematic literature review of software requirements prioritization research. *Information and software technology* 2014; 56(6): 568–585.

55. Khan JA, Khan SP, Qasim I, Khan YH. An Evaluation of Requirement Prioritization Techniques with ANP. *International Journal of Advanced Computer Science and Applications* 2016.

56. Carstens L, Toni F. Using argumentation to improve classification in natural language problems. *ACM Transactions on Internet Technology (TOIT)* 2017; 17(3): 1–23.

57. Carstens L, Toni F. Towards relation based argumentation mining. *Proceedings of the 2nd Workshop on Argumentation Mining* 2015: 29–34.

58. Baroni P, Romano M, Toni F, Aurisicchio M, Bertanza G. Automatic evaluation of design alternatives with quantitative argumentation. *Argument & Computation* 2015; 6(1): 24–49.

59. Achananuparp P, Lubis IN, Tian Y, Lo D, Lim EP. Observatory of trends in software related microblogs. *Proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering* 2012: 334–337.

60. Pagano D, Brügge B. User involvement in software evolution practice: a case study. *2013 35th International Conference on Software Engineering (ICSE)* 2013: 953–962.

61. Binkhonain M, Zhao L. A review of machine learning algorithms for identification and classification of non-functional requirements. *Expert Systems with Applications: X* 2019; 1: 100001.

62. Santos R, Groen EC, Villela K. An Overview of User Feedback Classification Approaches.. *REFSQ Workshops* 2019.

63. Hartmann J, Huppertz J, Schamp C, Heitmann M. Comparing automated text classification methods. *International Journal of Research in Marketing* 2019; 36(1): 20–38.

64. Sebastiani F. Machine learning in automated text categorization. *ACM computing surveys (CSUR)* 2002; 34(1): 1–47.

65. Maalej W, Nabil H. Bug report, feature request, or simply praise? On automatically classifying app reviews. *2015 IEEE 23rd International Requirements Engineering Conference, RE 2015 - Proceedings* 2015: 116–125.

66. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of machine Learning research* 2003; 3(Jan): 993–1022.

67. Chawla NV, Japkowicz N, Kotcz A. Special issue on learning from imbalanced data sets. *ACM SIGKDD explorations newsletter* 2004; 6(1): 1–6.

68. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 2002; 16: 321–357.

69. He H, Bai Y, Garcia EA, Li S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)* 2008: 1322–1328.

70. More A. Survey of resampling techniques for improving classification performance in unbalanced datasets. *arXiv preprint arXiv:1608.06048* 2016.

71. Kotsiantis S, Kanellopoulos D, Pintelas P, others . Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering* 2006; 30(1): 25–36.

72. Batista GE, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter* 2004; 6(1): 20–29.

73. Lippi M, Torroni P. Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology (TOIT)* 2016; 16(2): 1–25.

74. Dunne PE, Hunter A, McBurney P, Parsons S, Wooldridge M. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence* 2011; 175(2): 457–486.

75. Ferilli S, Pazienza A. An abstract argumentation-based approach to automatic extractive text summarization. *Italian Research Conference on Digital Libraries* 2018: 57–68.

76. Nielsen RD, Ward W, Martin JH. Recognizing entailment in intelligent tutoring systems. *Natural Language Engineering* 2009; 15(4): 479–501.

77. Khan FM, Khan JA, Assam M, Almasoud AS, Abdelmaboud A, Hamza MAM. A Comparative Systematic Analysis of Stakeholder's Identification Methods in Requirements Elicitation. *IEEE Access* 2022; 10: 30982–31011.

78. Zhang S, Wen L, Anwer S, Liu B. Using composition trees to matching software requirements—An external agency's approach to support software acquisition. *Software: Practice and Experience*.

□