Trust Evaluation Framework for Social Review Platforms

[1]Mr. P. Rajendra prasad
Associate Professor,Dept . Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: rajipe@gmail.com

[2]M. Lavanya
UG Student, Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: madigalavanya9550@gmail.com

[3]D. Vineetha
UG Student, Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: dornavineetha1234@gmail.com

[4]P. Pavani
UG Student, Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: pothanaboinapavani39@gmail.com

Abstract—Social Networks represent a cornerstone of our daily life, where the so-called social reviewing systems (SRSs) play a key role in our daily lives and are used to access data typically in the form of reviews. Due to their importance, social networks must be trustworthy and secure, so that their shared information can be used by the people without any concerns, and must be protected against possible attacks and misuses. One of the most critical attacks against the reputation system is represented by mendacious reviews. As this kind of attacks can be conducted by legitimate users of the network, a particularly powerful solution is to exploit trust management, by assigning a trust degree to users, so that people can weigh the gathered data based on such trust degrees. Trust management within the context of SRSs is particularly challenging, as determining incorrect behaviors is subjective and hard to be fully automatized. Several attempts in the current literature have been proposed; however, such an issue is still far from been completely resolved. In this study, we propose a solution against mendacious reviews that combines fuzzy logic and the theory of evidence by modeling trust management as a multi-criteria multi-expert decision making and exploiting the novel concept of time-dependent and content-dependent crown consensus. We empirically proved that our approach outperforms the main related works approaches, also in dealing with sockpuppet attacks.

Keywords: Social Networks, Social Reviewing Systems (SRSs), Trust Management, Reputation System, Mendacious Reviews, Trust Degree, Fuzzy Logic, Theory of Evidence, Time- dependent Consensus, Content-dependent Consensus, Crown Consensus, Sockpuppet Attacks.

I.INTRODUCTION

As well known, the online social networks are Internet-enabled applications used by people to establish social relations with the other individuals sharing similar personal interests and/or activities. Apart from exchanging per sonal data, such as photographs or videos, mainly all these applications allow their users to share comments and opinions on specific topics, so as to suggest objects or places of interest (e.g., TripAdvisor, Foursquare, etc.) or to provide social environments able to facilitate particular tasks (e.g., the search of a job as in LinkedIn, the answer to research questions as in ResearchGate, purchases on Amazon, etc.). Due to this comment/opinion sharing, these social applications, which we will refer to as social reviewing systems (SRSs) have been extensively used when people need to make daily decisions, increasing their popularity. As a concrete example, most of us access to a preferable SRS before choosing a restaurant or buying something so as to get reviews and feedback. People are progressively and symbiotically dependent on them as proved by the advanced opinion modeling and analysis, exploiting the impact of neighbors on user preferences or approaching the existing information overload in SRS, such as. For this reason, the trustworthiness of SRS is particularly important, and a key concern for effective opinion dynamics and trust propagation within a community of user. In fact, SRSs suffer from forged messages and camouflaged/fake users that are able to avoid individuals take the right decision. This may raise several issues about privacy and security, mainly due to the fact that several personal and sensitive information are shared, and leaked, throughout SRS and that a person may choose to hide its true self and intentions behind a totally false virtual identity or a Bot (short for software robots) may mimic human behavior in SRS. In addition, threats in SRS, such as data leaks, phishing bait, information tampering, and so on, are never limited to a given social actor, but spread across the network like an infection by obtaining victims among the friends of the infested actors. So, an SRS provider needs to provide proper protection means to guarantee its trustworthiness.

Some works in the current literature, such as, mostly deal only with forging messages as this can be easily resolved by using cryptography. However, the second kind of malicious behavior caused by camouflaged/fake users is still an open issue. During the last decade, several solutions have been proposed in order to deal with the problem of camouflaged/fake users. The issue of providing privacy has led to the adoption of access control means, while counteracting forging nodes/identities and social links/connections demanded authentication of users and exchanged messages. Mostly, such mechanisms aim at approaching external attackers or intruders, while thwarting legitimate participants in the SRS acting in a malicious way is extremely challenging. A naive way to protect against malicious individuals is to have users being careful when choosing with whom to have a relationship. Two users in social networks may have various kinds of relationships: 1) in Facebook-like systems users can indicate others as "friends," or 2) in Instagram-like systems a user can "follow" others.

II. LITERATURE SURVEY :

Haewon Byeon et al. (2024), in their paper *"Trustworthiness Text Detection Over Social Media Usage: A Supervised Sampling Approach for the Social Web of Things"*, propose a novel method for detecting trustworthiness in social media communication. The approach utilizes a Naïve Bayes Support Vector Machine (NBSVM) classifier within a supervised sampling framework. To enhance the evaluation of sample content, the authors apply Principal Component Analysis (PCA) and further optimize model parameters through an improved PCA-signed directed graph algorithm. This method aims to effectively identify trustworthiness-related patterns in the context of the Social Web of Things, addressing challenges in information reliability and communication quality.

Simran Chaudhry et al. (2024), in their research titled *"Trustworthiness Detection in Social Network Using Machine Learning Approach"*, investigate the use of Support Vector Machine (SVM) for trustworthiness classification within social networks. Their model is evaluated against a variety of parameters, highlighting the effectiveness of SVM in distinguishing between trustworthy and untrustworthy content. The study contributes to the field by showcasing how a conventional machine learning algorithm like SVM can be effectively applied in the domain of social network trust analysis, offering a promising baseline for future research enhancements.

Adewale A. Adewumi and Aderemi O. Adewumi (2021), in their comprehensive survey *"Advances in Trustworthiness Detection for Email Trustworthiness, Web Trustworthiness, Social Network Trustworthiness, and Review Trustworthiness"*, explore recent developments in trustworthiness detection across multiple digital domains. They review both machine learning models and nature-inspired computing approaches. The paper emphasizes the computational limitations of traditional machine learning algorithms and introduces nature-inspired models as a way to overcome these limitations. Their work provides a broad comparative analysis, laying a foundation for further advancements in hybrid and efficient trustworthiness detection methodologies.

III. METHODOLOGY

Identifying deceptive or "mendacious" reviews, a multicriteria decision-making approach is employed. This method evaluates reviews based on several criteria, and uniquely introduces the concepts of time-dependent and content-dependent crown consensus. This implies that the perceived quality of a review is not static, but evolves based on how a "crowd" or community perceives it over time, and also based on the review's actual content. This dynamic evaluation helps in flagging potentially dishonest feedback. Secondly, to assess user trustworthiness, the system utilizes reputation aggregation through the Dempster- Shafer (D-S) combination rule.

A. SYSTEM ARCHITECTURE

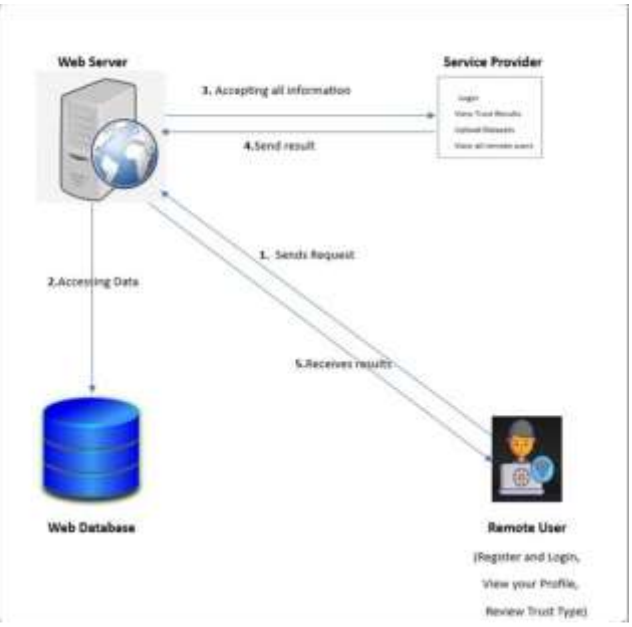


Fig: System Architecture

This architectural diagram illustrates a typical client-server interaction involving a "Remote User," a "Web Server," a "Web Database," and a "Service Provider." The Remote User initiates requests to the Web Server, which acts as the central hub. The Web Server is responsible for accessing data from the Web Database to fulfill user requests and also communicates with the Service Provider to accept information and send results. The Service Provider appears to house core business logic and potentially sensitive operations like login, viewing trust results, and managing datasets. Finally, the Web Server delivers the processed results back to the Remote User, completing the request-response cycle.

1. Remote User Sends Request:

The process begins with the "Remote User" initiating a request to the "Web Server." This request could be for various actions, such as logging in, viewing their profile, reviewing trust types, or performing other operations listed under the "Service Provider" (like viewing trust results, uploading datasets, or viewing all remote users).

2. Web Server Accessing Data:

Upon receiving the request, the "Web Server" interacts with the "Web Database." This step involves the Web Server retrieving or storing data relevant to the Remote User's request from the database. For instance, if the user is logging in, the Web Server would access the database to verify credentials.

3. Web Server Accepting All Information (from Service Provider):

After (or sometimes concurrently with) accessing its own database, the "Web Server" communicates with the "Service Provider." The arrow indicates that the Web Server is "Accepting all information" from the Service Provider. This suggests that the Service Provider holds the core logic and perhaps sensitive user data or functionalities that the Web Server needs to fulfil the Remote User's request. For example, the Service Provider might handle authentication, authorization, or the processing of trust- related information.

4. Web Server Sends Result (to Service Provider):

This step, where the "Web Server" "Send[s] result" to the "Service Provider," seems a bit unusual in a typical request-response flow unless the Web Server is acting as an intermediary that processes something and then reports back to the Service Provider, or perhaps it's sending data *to be processed* by the Service Provider. Given the overall flow, it's more likely that the Service Provider dictates the actions, and the Web Server might be sending back the outcome of its interaction with the Web Database, or perhaps forwarding information for the Service Provider to act upon.

5. Remote User Receives Results:

Finally, after all the internal processing between the Web Server, Web Database, and Service Provider, the "Web Server" sends the final "results" back to the "Remote User." This is the response to the initial request, such as a successful login, a displayed profile, or the outcome of a data upload.

IV. RESULT AND ANALYSIS

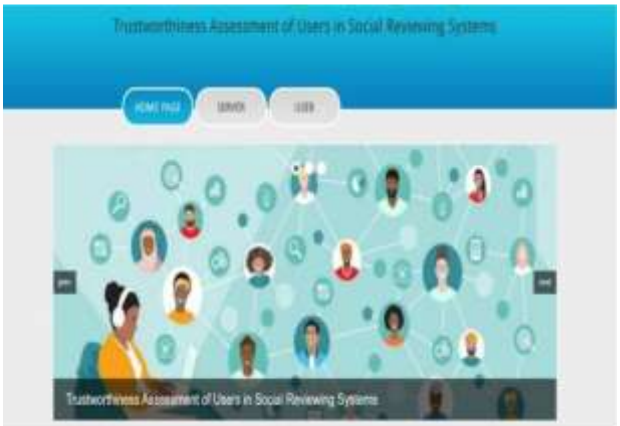

Trustworthiness Assessment of Users in Social Reviewing Systems
HOME PAGE
SERVER
USER
Trustworthiness Assessment of Users in Social Reviewing Systems

Fig-1: Home Page

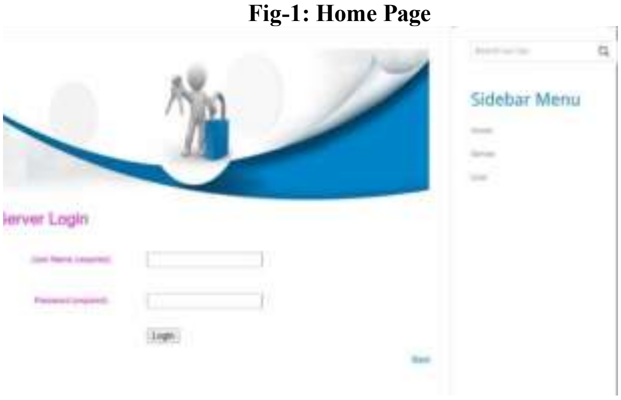


Fig-2: Server Login Page

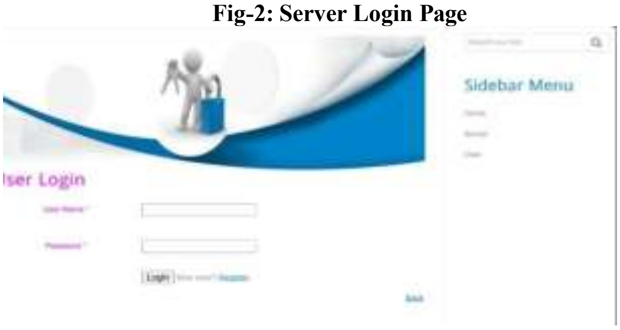


Fig-3: User Login

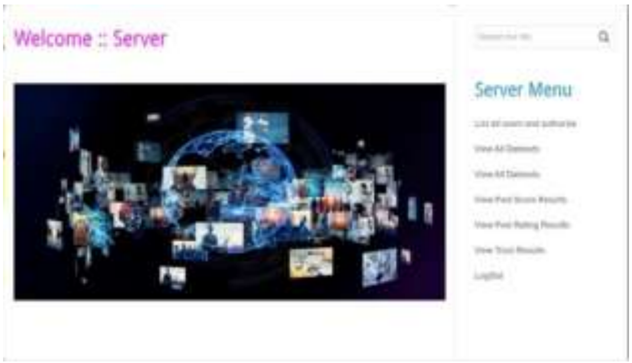


Fig-4: Server Menu

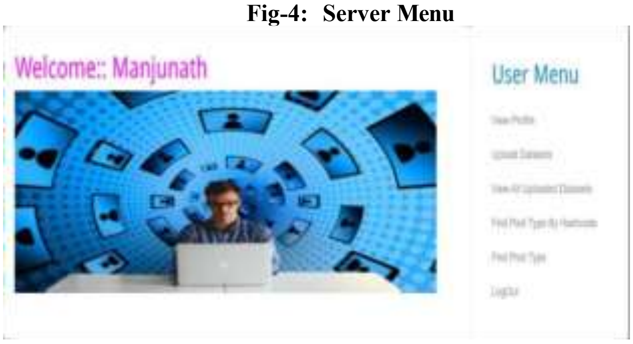
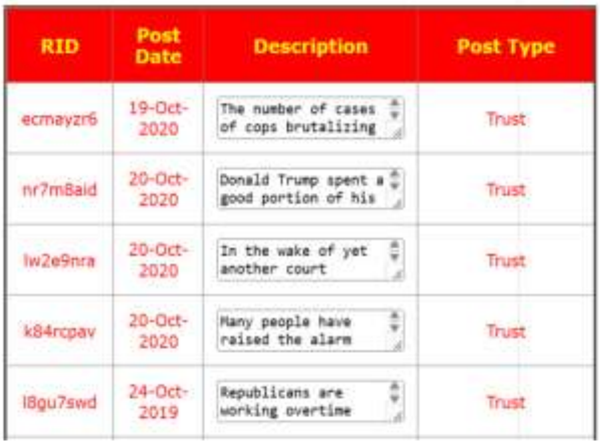
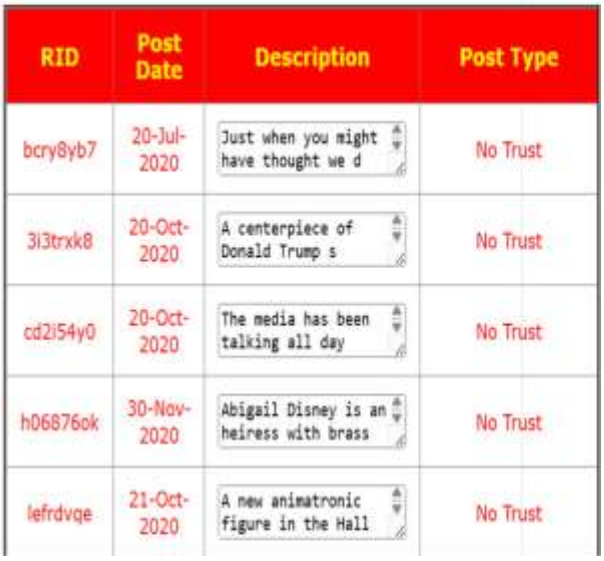


Fig-5: User Menu



RID	Post Date	Description	Post Type
ecmayzr6	19-Oct-2020	The number of cases of cops brutalizing	Trust
nr7m8aid	20-Oct-2020	Donald Trump spent a good portion of his	Trust
lw2e9nra	20-Oct-2020	In the wake of yet another court	Trust
k84rcpav	20-Oct-2020	Many people have raised the alarm	Trust
l8gu7swd	24-Oct-2019	Republicans are working overtime	Trust

Fig-6: Trust Results



RID	Post Date	Description	Post Type
bcry8yb7	20-Jul-2020	Just when you might have thought we d	No Trust
3i3trxk8	20-Oct-2020	A centerpiece of Donald Trump s	No Trust
cd2i54y0	20-Oct-2020	The media has been talking all day	No Trust
h06876ok	30-Nov-2020	Abigail Disney is an heiress with brass	No Trust
lefrdvqe	21-Oct-2020	A new animatronic figure in the Hall	No Trust

Fig-7: No trust Result

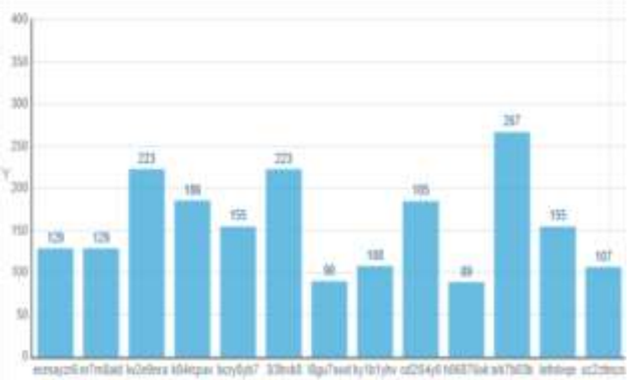
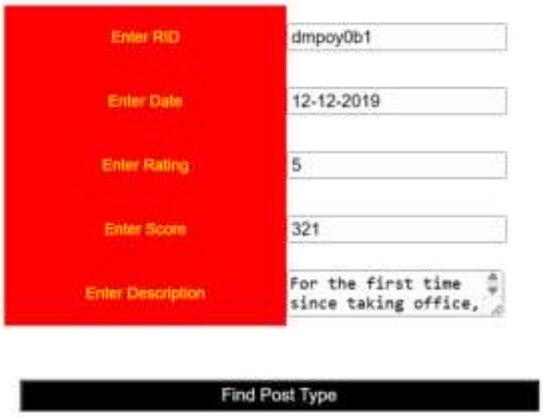
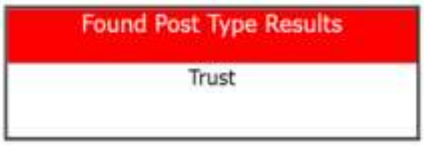


Fig-8: Trust score of each user



Enter RID	dmpoy0b1
Enter Date	12-12-2019
Enter Rating	5
Enter Score	321
Enter Description	For the first time since taking office,

Fig-9: Finding post type



Found Post Type Results
Trust

Fig-10: Result of the post V.CONCLUSION

This project presents a novel approach to addressing the critical issue of trustworthiness in social reviewing systems by integrating advanced techniques for both mendacious review identification and user trustworthiness assessment. By employing a multicriteria decision-making framework with innovative time-dependent and content-dependent crown consensus measures, the system can more accurately identify deceptive reviews, moving beyond static, one-dimensional evaluations. Furthermore, the application of the Dempster-Shafer theory for aggregating diverse reputation scores provides a robust and theoretically sound method for inferring user trustworthiness, effectively handling uncertainty and conflicting evidence. The proposed system offers a comprehensive solution to enhance the reliability of online reviews, thereby empowering consumers to make more informed decisions and fostering a healthier, more credible online review ecosystem. The combination of these methodologies contributes significantly to the body of knowledge in online trust and reputation management, offering practical implications for platform administrators seeking to combat review manipulation effectively.

VI. FUTURE SCOPE

The proposed project lays a strong foundation for future research and development in user trustworthiness assessment in social reviewing systems. Several promising avenues can be explored:

Real-time Trustworthiness Assessment: Develop mechanisms for real-time or near real-time assessment of reviews and user trustworthiness, which is crucial for dynamic social reviewing environments. This would involve optimizing the processing pipeline and potentially leveraging stream processing technologies. Explainability and Interpretability: Enhance the explainability of the trustworthiness assessment results. Provide insights into why a review is flagged as mendacious or how a user's trustworthiness score was derived, helping platform administrators understand and act upon the system's outputs. This could involve visualizing criteria contributions or D-S evidence.

Active Learning and User Feedback Integration: Incorporate user feedback on the trustworthiness predictions (e.g., flagging reviews as fake, reporting users) into an active learning framework. This would allow the model to continuously learn and improve its performance based on human intelligence. Cross-Platform Trust Transfer: Investigate methods for transferring trustworthiness assessments across different social reviewing platforms. A user's reputation on one platform might inform their trustworthiness on another, potentially using federated learning or distributed ledger technologies.

VII. REFERENCES

[1] Byeon, H., Jha, S., Keshta, I., Bhatt, M. W., Singh, P. P., Jindal, L., & Vijaya Lakshmi, T. R. (2024). Spam Text Detection Over Social Media Usage: A Supervised Sampling Approach for the Social Web of Things. *IEEE Systems, Man, and Cybernetics Magazine, 10*(2), 32–39. https://doi.org/10.1109/msmc.2023.3343950

[2] Chaudhry, S., Dhawan, S., & Tanwar, R. (2020). Spam Detection in Social Network Using Machine Learning Approach. *International Journal of Engineering Sciences & Research Technology, 9*(5), 114–123. https://www.ijesr.org/admin/uploads/Article%20%20pp%201 14-123.pdf

[3] Adewumi, A. A., & Adewumi, A. O. (2021). Advances in Trustworthiness Detection for Email Trustworthiness, Web Trustworthiness, Social Network Trustworthiness, and Review Trustworthiness.