IJAREEIE

e-ISSN: 2278-8875
p-ISSN: 2320-3765

International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering

Volume 13, Issue 4, April 2024

Impact Factor: 8.317

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 4, April 2024||

| DOI:10.15662/IJAREEIE.2024.1304065 |

Online Fake Review Detector Using Arm Processor

Bindu M, Gautham T M, Manikanta S, Pavan Raj S, Mr. Harish S V

UG Students, Department of Electronics and Communication, NIE Institute of Technology Mysuru, Karnataka, India

Assistant Professor, Department of Electronics and Communication, NIE Institute of Technology Mysuru, Karnataka, India

ABSTRACT: In the last few years review sites are more and more confronted to spread of misinformation, to promote or to damage certain businesses various opinion spam's are done either to mislead the human readers or the sentiment analysis or opinion mining systems which are automated. In the last few years because of this reason various approaches have been proposed so that the credibility of the user generated content can be assessed. The analysis of the main review and the reviewer-centric features are proposed to detect the fake reviews by using supervised machine learning approaches rather than the unsupervised approaches which are based on graphical methods. In e-commerce, user reviews can play a significant role in determining the revenue of an organization. Online users rely on reviews before making decisions about any product and service. As such, the credibility of online reviews is crucial for businesses and can directly affect companies' reputation and profitability. That is why some businesses are paying spammers to post fake reviews. These fake reviews exploit consumer purchasing decisions. Consequently, the techniques for detecting fake reviews have extensively been explored in the past twelve years. However, there still lacks a survey that can analyses and summarize the existing approaches. To bridge up the issue, this survey paper details the task of fake review detection, summing up the existing datasets and their collection methods. It analyses the existing feature extraction techniques.

KEYWORDS: Fake Review ,Automatic , technology, Companies , Online, Detector etc..,

.

I.INTRODUCTION

The user generated content is increasing popularity on the social websites without any form of trusted external control and thus there are no means to verify which content generated by the user is believable or which source is reliable. The consequences of spread of such misinformation are negative and it causes harm to user as well as businesses. The different subset of characteristics i.e. features often considered by various approaches connected to both reviews and reviewers as well as to the network structure linking distinct entities on the review-site in exam. The main purpose is to provide analysis of the main review and review -centric features that have been proposed to detect fake reviews, in particular approaches that employ supervised machine learning techniques. Fake reviews, fake comments, fake blogs, fake social networking postings, deceptive messages are identified by opinion spam detection. The review-centric sites such as yelp can be considered while detecting fake review detection. Unsupervised approaches have been incorporated so far for detecting fake reviews which are based on graphical methods but are not much reliable. The supervised techniques consider distinct features generated from the reviews as well as the behavior of the reviewer. Nowadays, when clients choose to draw a choice about offerings or products, critiques end up the predominant supply of their information. For example, when clients take the initiation to e book a hotel, they examine the opinions on the opinions of different clients on the services.

II. SYSTEM MODEL AND ASSUMPTIONS

The processor is complemented by two Programmable Real-Time Units (PRUs), which provide real-time processing capabilities for applications requiring precise timing and control. The Beagle Bone's architecture includes expansion headers (P8 and P9) that expose GPIO, I2C, SPI, UART, and other interfaces for connecting to a variety of devices. Its pin diagram, spread across these headers, provides a visual representation of the available input/output pins. For example, P8 and P9 headers contain pins for power supply, communication interfaces, analog inputs, and more. Understanding the pin diagram is crucial for hardware interfacing, enabling developers to connect sensors, actuators,

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 4, April 2024||

| DOI:10.15662/IJAREEIE.2024.1304065 |

and other components to create a diverse range of projects, from IoT applications to embedded systems. Black is a low-cost, high-expansion, community-supported development platform for developers and hobbyists. Users can boot Linux in under 10 seconds and get started on development in less than 5 minutes with just a single USB cable. Like its predecessors, the Beagle Bone Black is designed to address the Open Source Community, early adopters, and anyone interested in a low cost ARM® Cortex™-A8 based processor. It has been equipped with a minimum set of peripherals to allow the user to experience the power of the processor and also offers access to many of the interfaces and allows for the use of add-on boards called capes, to add many different combinations of features. A user may also develop their own board or add their own circuitry.

- **Python for Machine Learning:**
  Python has emerged as a go-to programming language for machine learning tasks due to its simplicity, readability, and a rich ecosystem of libraries. Leveraging libraries like scikit-learn, TensorFlow, and Natural Language Toolkit (NLTK), we can efficiently process and analyze textual data, a key component in fake review detection.

- **Natural Language Processing (NLP) for Review Analysis:**
  To distinguish between genuine and fake reviews, the system must comprehend the underlying sentiments and contexts in the text. NLP techniques play a vital role in achieving this understanding. By employing tokenization, stemming, and sentiment analysis, the system can extract meaningful features from reviews, providing a basis for classification.

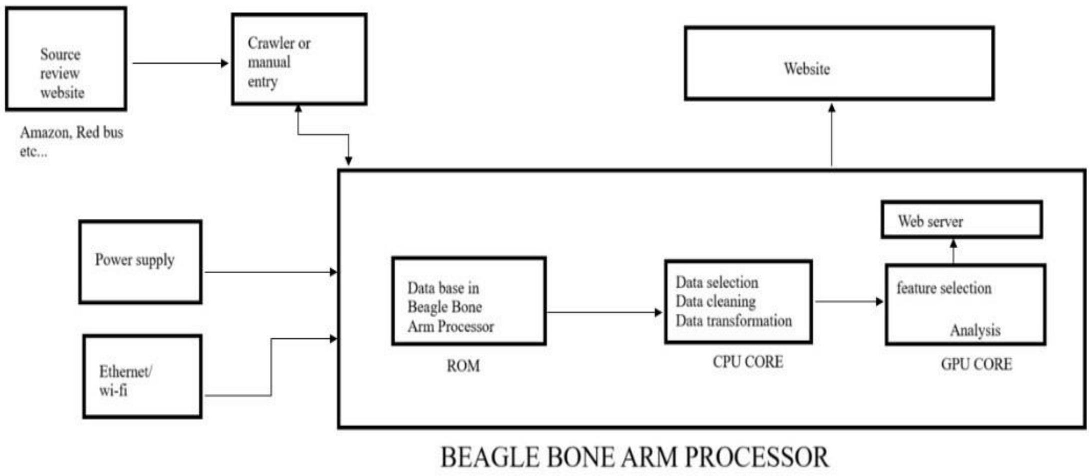


Fig. 1 System Model

Crawler/Manual Entry The first step is to collect data containing reviews. This can be done through web scraping (crawler) from various review platforms or through manual entry. Both methods aim to gather a diverse dataset that reflects different products and services.

Data Selection After collecting raw data, the next step is to filter and select relevant information for analysis. This includes extracting review text, ratings, product details, and other metadata. The goal is to create a focused dataset that represents the target domain of interest.

Data Cleaning the data is crucial to ensure its quality and reliability. This involves removing duplicates, handling missing values, correcting errors, and standardizing formats. Clean data is essential for accurate analysis and model training.

Data Transformation Transforming the data involves converting raw text into a format suitable for analysis. This may include text preprocessing steps such as tokenization, stemming, and removing stop words. Additionally, numerical features like ratings may be normalized to a consistent scale for better comparison.

Feature Selection Feature selection is a critical step in building a machine learning model. It involves choosing the most

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 4, April 2024||

| DOI:10.15662/IJAREEIE.2024.1304065 |

relevant features that contribute to the detection of fake reviews. Features can include sentiment analysis scores, word frequency, and other linguistic or contextual indicators.
Analysis phase involves applying machine learning algorithms to the selected features to train a model for fake review detection. Common algorithms include natural language processing (NLP) techniques and supervised learning models. The model learns patterns from the labeled data to distinguish between genuine and fake reviews.

Web Server The web server component facilitates the deployment and interaction with the detection model. It provides an interface for users to input reviews, and it delivers the model's predictions in real-time. This allows users to quickly assess the authenticity of reviews and make informed decisions based on the results.

III.METHODOLOGY

A. Review-centric Features The first class of features that have been considered is constituted by those related to a review. They can be extracted both from the text constituting the review, i.e., textual features, and from meta-data connected to a review, i.e., meta-data features. A large part of reviews are singletons, i.e., there is only one review written by a given reviewer in a certain period of time for this kind of reviews, specific features must be designed.

- Textual Features
- Meta-data Features:

B. Reviewer-centric features This group of features is composed of features related to the reviewer's behaviour. In this way it is possible to go beyond the content and meta-data associated with a review, which are limited for classification, and considering the behaviour of users in general in writing reviews. Textual features Rating features Temporal features

C. Choice of the classifier and implementation The majority of supervised classifiers to tackle the issue of opinion spam detection are based on Naıve Bayes or Support Vector Machines (SVM). To implement the classifier, the Python programming language has been employed, as it is used by a large community of developers, thus offering a vast set of tools and libraries for different aims.

D. Choice of the dataset The classification provided by Yelp has been used as a ground truth, where recommended reviews correspond to 'genuine' reviews, and not recommended reviews correspond to 'fake' ones. The strengths of these datasets are The high number of reviews per user, which allows to consider the behavioural features of each user. The diversified kinds of entities reviewed, i.e., restaurants and hotels .The datasets only contain basic information, such as the content, label, rating, and date of each review, connected to the user who generated them.

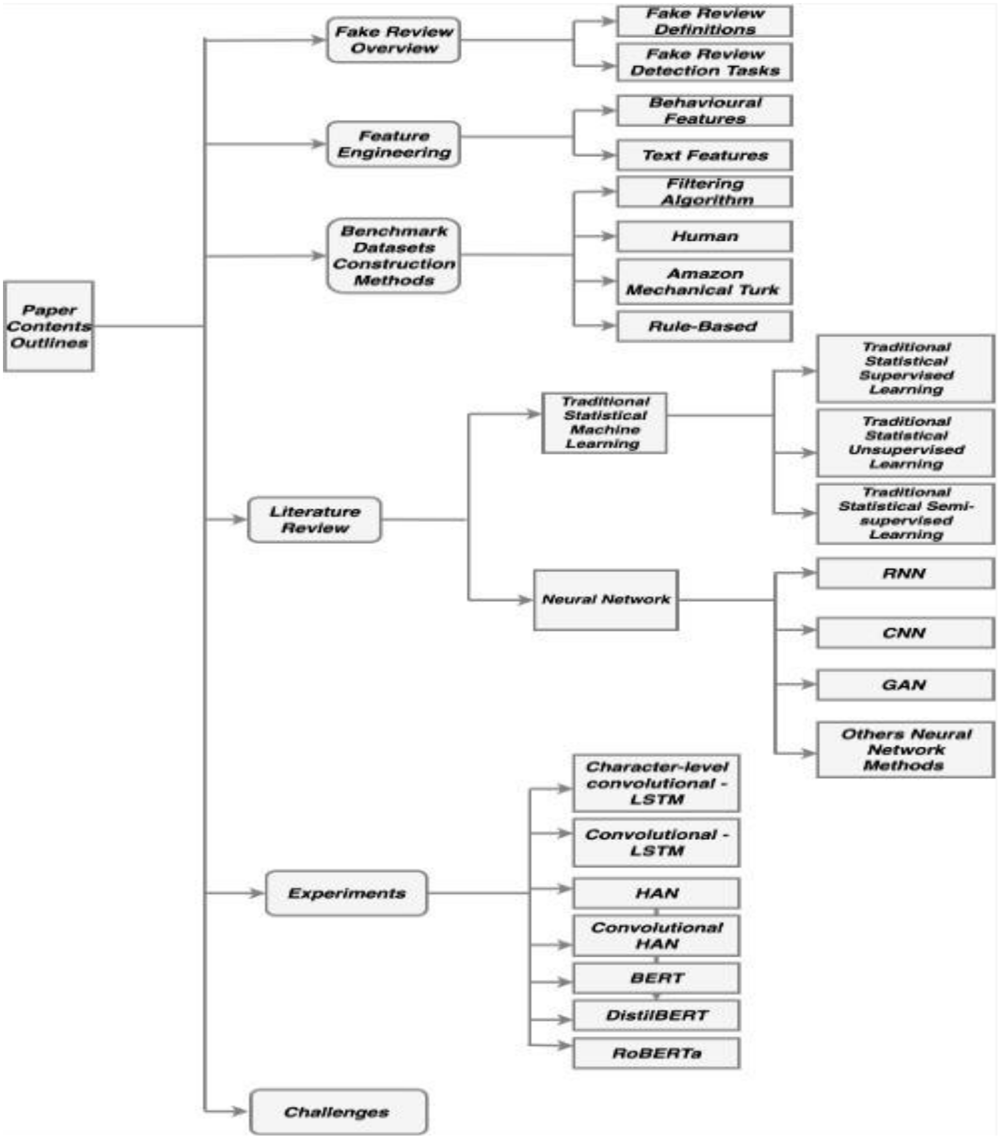
E. Balancing data Imbalanced data represents one of the major issues that have to be tackled when performing supervised classification. In the training phase, if the unbalancing of training data is not considered, there is the risk that the classifier learns mainly from the largest class of labelled data therefore neglecting the minority class. The oversampling method is considered, it consists in augmenting the minority class to balance it with the largest one.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 4, April 2024||

| DOI:10.15662/IJAREEIE.2024.1304065 |

IV.FLOW CHART



V.SURVEY DESCRIPTION

This survey aims to collect feedback on several aspects related to online fake review detection, with a focus on utilizing ARM processors for efficient processing. Your responses will aid in understanding user requirements, preferences, and potential challenges in implementing such a system.

Demographic Information: Understand your background and experience related to online reviews and ARM processors. Understanding of Fake Reviews: Assess your familiarity with fake reviews, including detection techniques and challenges.

ARM Processor Knowledge: Evaluate your knowledge and experience with ARM processors and their applications. Features and Functionality: Gather insights on the desired features and functionality of an online fake review detection system powered by ARM processors.

Challenges and Concerns: Identify potential challenges, concerns, and limitations in implementing such a system.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 4, April 2024||

| DOI:10.15662/IJAREEIE.2024.1304065 |

Suggestions and Feedback: Provide any additional suggestions or feedback you may have regarding the development of this system.

VI.RESULT

Pattern Recognition: Fake review detectors utilize pattern recognition techniques to identify anomalies in review data. These patterns may include linguistic cues, review behaviours, and temporal factors that deviate from the norm of genuine reviews.

Machine Learning: Many detectors employ machine learning algorithms trained on labelled datasets of genuine and fake reviews. These algorithms learn to recognize patterns and features indicative of fake reviews, allowing them to make predictions on unseen data.

Natural Language Processing (NLP): NLP techniques are applied to analyse the language and sentiment of reviews. Fake reviews often exhibit distinct linguistic characteristics, such as excessive use of superlatives, unusual vocabulary, or grammatical errors.

Reviewer Behaviour Analysis: By examining the behaviour of reviewers, detectors can identify suspicious patterns such as posting frequency, review length, and the distribution of ratings. For example, a sudden influx of reviews from new or inactive accounts may raise red flags.

Metadata Analysis: Review metadata, including timestamps, IP addresses, and device information, can provide valuable insights into the authenticity of reviews. Clusters of reviews originating from the same IP address or device may indicate coordinated efforts to manipulate ratings.

Cross-Platform Analysis: Detecting fake reviews often involves analysing data from multiple platforms. Fake reviewers may target several platforms simultaneously, leaving behind similar patterns of behaviour across different sites.

Temporal Analysis: Temporal factors, such as the timing and frequency of reviews, are also considered. Anomalous patterns, such as a sudden spike in reviews following a product launch or a negative event, may suggest manipulation.

Ensemble Methods: Some detectors use ensemble methods, combining the outputs of multiple algorithms or models to improve accuracy and robustness. This approach helps mitigate the limitations of individual methods and enhances overall detection performance.

Feedback Loop: Continuous monitoring and feedback mechanisms are crucial for improving detection algorithms over time. Feedback from users, manual review processes, and updates to detection algorithms help adapt to evolving tactics used by perpetrators of fake reviews.

VII.FUTURE SCOPE AND DISCUSSION

Advanced Machine Learning Techniques Continued advancements in machine learning algorithms, such as deep learning and reinforcement learning, could improve the accuracy and efficiency of fake review detection systems.

Semantic Analysis Future detectors may incorporate more sophisticated semantic analysis techniques to better understand the meaning and context of reviews, enabling them to identify subtle patterns indicative of fake reviews.

Multimodal Analysis Integrating multiple data modalities, such as text, images, and audio, could provide richer information for fake review detection, particularly in platforms where reviews contain multimedia content.

Behavioural Biometrics Leveraging behavioural biometrics, such as typing patterns, mouse movements, and browsing behaviour, may offer additional signals for identifying fake reviewers and enhancing detection accuracy.

Blockchain Technology The use of blockchain technology could provide a decentralized and tamper-proof framework for storing and verifying review data, reducing the risk of manipulation and enhancing trust in online reviews.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 4, April 2024||

| DOI:10.15662/IJAREEIE.2024.1304065 |

Explainable AI Developing explainable AI techniques will be crucial for enhancing transparency and trust in fake review detection systems, allowing users to understand the rationale behind classification decisions.

Real-time Detection Improvements in real-time detection capabilities will enable platforms to swiftly identify and mitigate fraudulent activities, minimizing their impact on users and businesses.

Cross-domain Detection Expanding fake review detection beyond e-commerce platforms to other domains such as hospitality, healthcare, and education will be important for addressing fraudulent activities across various sectors.

User-Centric Solutions Incorporating user feedback and preferences into detection algorithms can enhance the user experience and ensure that detection methods align with users' expectations and requirements.

VIII.CONCLUSION

Online fake review detectors employ a combination of pattern recognition, machine learning, natural language processing, reviewer behaviour analysis, metadata analysis, cross-platform analysis, temporal analysis, ensemble methods, and feedback mechanisms to distinguish between genuine and fraudulent reviews. By leveraging these theories and methodologies, detectors can effectively identify anomalies indicative of fake reviews, thereby enhancing the reliability and integrity of online review platforms. Ultimately, the deployment of robust detection systems contributes to fostering consumer trust, facilitating fair competition among businesses, protecting brand reputation, and promoting transparency in the online marketplace.

ACKNOWLEDGMENT

This work was supported by our project guide, Mr. Harish S V, Assistant Professor, ECE, NIEIT. We are thankful to the guide and faculties of our college who helped us in proposing this system

REFERENCES

[1] R. Barbado, O. Araque, and C. A. Iglesias, “A framework for fake review detection in online consumer electronics retailers,” Information Processing & Management, vol. 56, no. 4, pp. 1234 – 1244, 2019

[2] M. J. H. Mughal, “Data mining: Web data mining techniques, tools and algorithms: An overview,” Information Retrieval, vol. 9, no. 6, 2018.

[3] E. Elmurngi and A. Gherbi, Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques. IARIA/DATA ANA- LYTICS, 2017.

[4] Julian Fontanarava, Gabriella Pasi, and Macro Viviani, “Feature Analysis for Fake Review Detection through Supervised Classification,” in 2017 International Conference on Data Science and Advanced Analytics. IEEE, 2017, pp. 658–666.

[5] F. Zhu and X. Zhang, “Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics,” Journal of marketing, vol. 74, no. 2, pp. 133–148, 2010.

[6] S. KC and A. Mukherjee, “On the temporal dynamics of opinion spamming: Case studies on yelp,” in Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2016, pp. 369–379.