International Journal of Industrial Engineering
2024;8(2):66-73.
ISSN: 2456-8449

Integrating ClaimBuster with Cloud Computing for Automated Fake News Detection

Thirumoorthy Arumugam[1*], P. Senthilraja[1], A. Gnanabaskaran[1]

[1]*Department of Computer Science and Engineering, KS Rangasamy College of Technology, Tiruchengode, Tamil Nadu, India.*

[*]*Corresponding author: thirumoorthy.arumugam@gmail.com*

Abstract. The rapid spread of fake news across digital platforms poses a serious threat to information integrity and public trust. Traditional fact-checking approaches are often limited by scalability and timeliness. This paper presents the integration of ClaimBuster, an automated fact-checking system, with cloud computing infrastructure to provide a scalable, accurate, and efficient solution for real-time misinformation detection. The proposed system integrates machine learning algorithms and Natural Language Processing (NLP) techniques to classify claims as true or false, supported by cloud-enabled elastic computing resources for handling high-volume, high-velocity data streams. Experimental evaluation demonstrates strong performance in terms of accuracy, precision, recall, and F1-score, with low false positive rates, confirming its robustness in real-world applications. Furthermore, advanced visualization tools, gamification features, and partnerships with media, academic, and governmental institutions enhance user engagement and the overall impact of the system. By prioritizing scalability, adaptability, privacy, and ethics, ClaimBuster within cloud computing offers a comprehensive framework for combating the global challenge of disinformation.

Keywords: Fake News Detection, ClaimBuster, Cloud Computing, Scalable Systems, Misinformation.

INTRODUCTION

Fake news poses a significant challenge by spreading disinformation and undermining trust in credible sources. While traditional fact-checking methods are effective, they are often time-consuming and are unable to keep pace with the sheer volume of digital content. ClaimBuster, a machine learning-based fact-checking tool, automates the detection of false claims, offering a scalable, precise, and efficient solution. By leveraging cloud computing infrastructure, ClaimBuster can analyze massive datasets in real time, identifying and assessing potentially misleading information across digital platforms. The system utilizes machine learning algorithms to assign likelihood scores to textual assertions, indicating their probability of being false. Cloud integration enhances ClaimBuster's efficiency, enabling it to handle high-velocity, high-volume data streams effectively. This approach provides media organizations, governments, and digital platforms with reliable, timely insights to combat the rapid dissemination of disinformation in today's digital ecosystems.

Section 2 addresses the challenges of fake news detection, including the rapid proliferation of online content and the inherent limitations of manual fact-checking. The section emphasizes the advantages of machine learning models, particularly ClaimBuster, combined with cloud computing to achieve scalability and high-volume data processing. Section 3 outlines the deployment of ClaimBuster on cloud platforms, detailing key processes such as data ingestion, model deployment, and real-time claim analysis. It further examines the integration of machine learning pipelines with cloud services, including distributed storage and processing capabilities. The section also highlights strategies for continuous learning and dataset expansion, which enhance the system's predictive accuracy and adaptability over time. Section 4 evaluates the performance of the system using metrics such as accuracy, precision, recall, and scalability. Case studies demonstrate ClaimBuster's ability to detect false claims in real time across high-volume data streams, highlighting its operational effectiveness and practical applicability in combating disinformation. Section 5 concludes with a comprehensive review of ClaimBuster and its integration with cloud computing for scalable and accurate fake news detection. Future enhancements are proposed to expand the system's reach and utility, including multilingual capabilities, real-time sentiment analysis, and integration with social media platforms.

Recent advancements in automated claim detection and fact-checking leverage enhanced language models and hybrid systems to improve reliability and efficiency. One approach employs LoRA-enhanced Large Language

Received: 13.06.2024 Revised: 15.08.2024 Accepted: 25.08.2024

Models (LLMs), which are fine-tuned using transfer learning and optimized computational frameworks to detect claims, claimers, and claim objects across diverse datasets [1]. Another method, VERITAS-NLI, integrates Natural Language Inference (NLI) with web scraping to automate the extraction of trustworthy information, enabling systematic evaluation of data credibility from online sources [2]. Comparative studies indicate that fine-tuned transformer models perform exceptionally well in fact-checking tasks, often rivaling larger language models [3]. Additionally, QuanTemp provides an open-domain benchmark for assessing numerical fact-checking systems, enhancing AI capabilities in processing quantitative statements and thereby strengthening automated verification applications [4].

Recent research on misinformation detection emphasizes multi-modal and domain-specific strategies to enhance accuracy and reach. ViMGuard introduces a multi-modal system for identifying deceptive video content, combining visual and textual cues to detect misinformation in real time, with applications in social media monitoring [5]. NumTEMP provides a benchmark for evaluating fact-checking systems that process historical and current statistical data, supporting AI models in validating quantitative claims effectively [6]. In the healthcare domain, specialized architectures have been developed to detect false information in medical records, leveraging machine learning algorithms to mitigate the spread of inaccurate health claims [7]. Additionally, studies in rural India highlight disparities in fact-checking accessibility and awareness, offering insights into improving information verification in marginalized and geographically isolated regions [8].

Accurate data validation is critical across domains such as digital journalism, public health, and finance. Addressing the growing need for automated and scalable solutions, recent systems in data-driven disinformation management are designed to handle massive datasets efficiently [9]. A comprehensive survey on automatic credibility assessment using LLMs highlights the potential of large language models to evaluate textual content in real time, providing an effective tool for automated credibility verification [10]. Targeted strategies for tracing the origins of misinformation examine specific websites and URLs, enabling news organizations and social media platforms to identify sources early and limit the spread of false information [11]. In the healthcare context, studies on combating false health claims on social media explore IT-based methods for detecting and mitigating misinformation, supporting public health monitoring and evidence-based policymaking [12].

Fact-checking automation faces technical limitations that can hinder effective misinformation management. A technographic review of past and present approaches examines these constraints, highlighting applications in public policy, social media, and journalism where rapid verification is critical [13]. Audio content verification presents unique challenges, as explored in studies auditing podcast claim credibility, which provide insights into effective fact-checking strategies for non-textual media [14]. In collaborative knowledge platforms, citation detection on Wikipedia supports fact-checking by identifying sentences that lack references, ensuring open-access information remains accurate and trustworthy [15]. Additionally, the evaluation of automated question generation for fact-checking demonstrates its potential to improve verification processes through structured question-answering frameworks, benefiting educational institutions, online news outlets, and legal organizations [16].

The use of Large Language Models (LLMs) for automated fact-checking presents both opportunities and challenges. A survey of LLM applications highlights their potential in building robust fact-checking infrastructures, emphasizing the need for fine-tuning these models to optimize misinformation detection [17]. To address sustainability concerns, research on energy-efficient model training using active learning explores methods that reduce computational power requirements while maintaining model performance, providing insights into environmentally conscious AI development [18]. Practical tools, such as browser extensions for misinformation detection, allow users to verify facts in real time, fostering critical engagement with online content and supporting proactive disinformation control. In parallel, educational approaches for teaching misinformation literacy equip learners with the skills to navigate complex digital media landscapes responsibly, promoting informed information consumption across various educational levels.

PROPOSED METHODOLOGY

Fake news has a profound impact on society, the economy, and politics worldwide. In the digital era, misinformation spreads rapidly, amplifying the influence of false information. Leveraging cloud computing allows scalable storage and processing of large datasets, enabling rapid detection and response to emerging misinformation trends. Integrating ClaimBuster within a robust cloud architecture equips the system to address fake news on a global scale. As illustrated in Figure 1, the cloud-based architecture comprises three main components: Data

Preprocessing, Fake News Detection Model, and Result Aggregation. The Data Preprocessing module cleans and organizes claim data to ensure quality input for analysis, while the Fake News Detection Model employs machine learning to classify claims as true or false based on the preprocessed data.

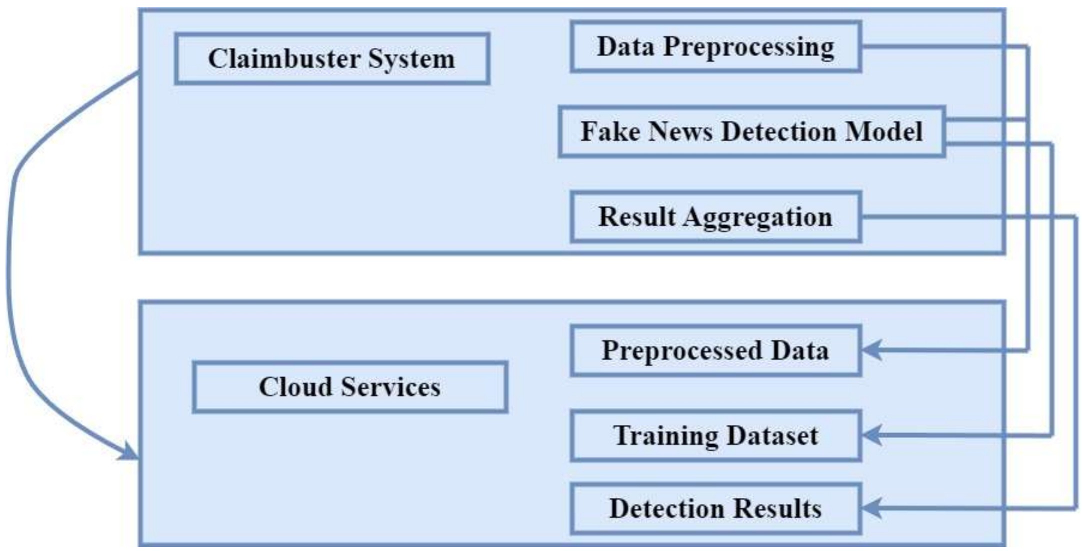

FIGURE 1. Block Diagram of ClaimBuster Cloud Architecture

Integrating ClaimBuster into a cloud computing environment is facilitated by its scalable system architecture. Diverse textual data sources, including news articles, social media posts, and API feeds, are ingested into the cloud platform for analysis. Leveraging cloud scalability, the system can efficiently process large volumes of data and detect fake news in real time. As depicted in Figure 2, the block diagram illustrates the integration of the ClaimBuster model within the cloud architecture, highlighting the Claim Processing Pipeline. Key components—Data Ingestion, Feature Extraction, and the Machine Learning Model—work together to identify and classify false claims effectively.

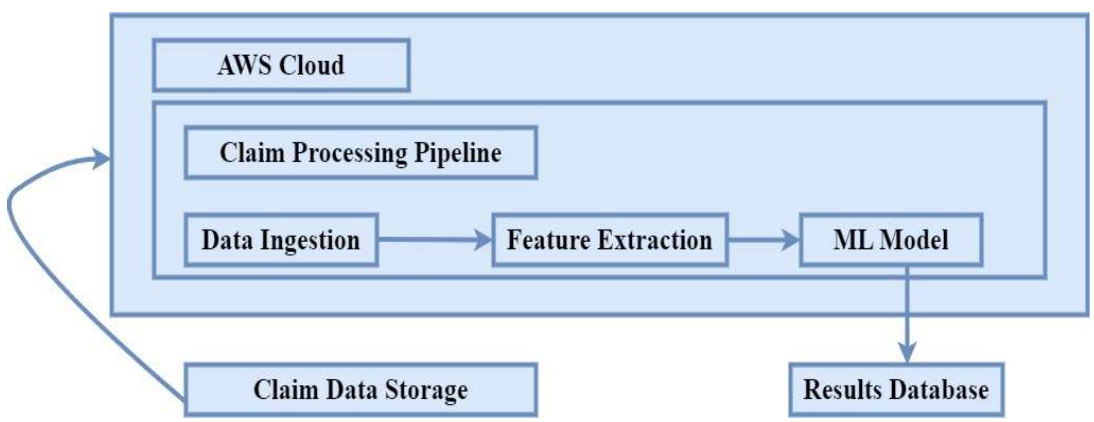

FIGURE 2. Block Diagram of ClaimBuster Model Integration in Cloud

The proposed system is built on a cloud platform, enhancing both reliability and functionality. To manage surges in data volume during news cycles or major events, it utilizes elastic computing resources that scale dynamically. As shown in Figure 3, the Claim Verification Workflow outlines the ClaimBuster verification process. Users submit claims through a User Query, which are processed by the Cloud Processing Unit, the system's central processor. The unit retrieves relevant fact-checking data from the Fact-Check Database and forwards it to the Validation Engine. Using machine learning techniques, the Validation Engine evaluates the claim against the collected data and delivers a result indicating whether the claim is true or false.

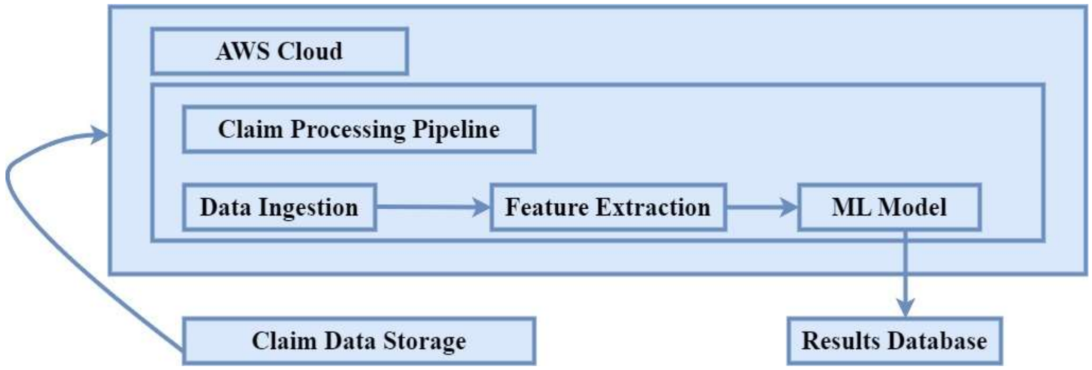

FIGURE 3. Block Diagram of Claim Verification Workflow

Precision is a key aspect of the proposed system's architecture. ClaimBuster leverages large datasets to train its machine learning models, enabling the identification and evaluation of claims across diverse contexts. Regular updates to these datasets ensure the system remains effective against emerging misinformation. Continuous monitoring and retraining on the cloud platform allow the system to adapt to new trends and evolving linguistic patterns. As depicted in the Operational Workflow in Figure 4, the ClaimBuster process begins with user-submitted claims, which are pre-processed for analysis. Invalid claims are promptly filtered out, while legitimate claims are compared against a fact database. Advanced machine learning techniques are then applied to detect and verify potentially false news.

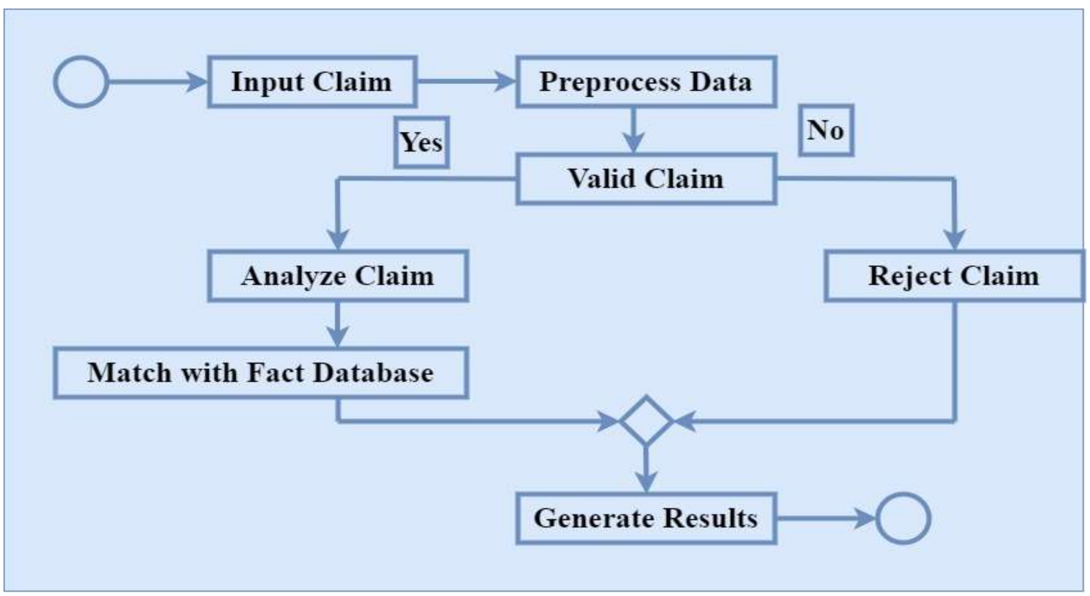

FIGURE 4. Workflow Diagram of ClaimBuster Operational Workflow

The deployment of a fake news detection system raises significant privacy and ethical considerations. Protecting the privacy of data subjects is paramount, and robust encryption methods secure data both at rest and in transit, in compliance with GDPR. Access control ensures that only authorized personnel can view sensitive information, maintaining confidentiality. Transparency is equally important, as users can comprehend the claim verification process through interpretable operations. To address ethical concerns such as algorithmic bias, system performance is continuously evaluated, and training datasets are diversified. By prioritizing privacy and ethics, the system fosters user trust and confidence. Figure 5 illustrates the ClaimBuster Cloud Ecosystem Overview, highlighting Data Sources, Cloud Services, and End-User Interfaces. Data sources supply claims, which the cloud infrastructure processes, while cloud services—including data storage, machine learning, and computational resources—enable scalable claim management.

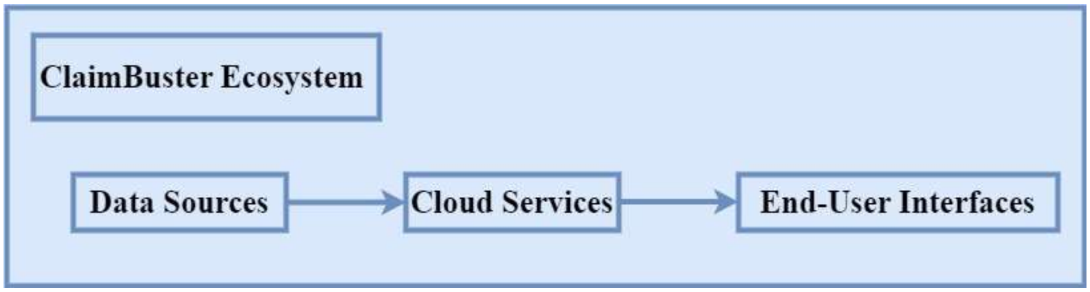

FIGURE 5. Overview Diagram of ClaimBuster in Cloud Ecosystem

The key features of ClaimBuster within cloud computing for detecting false information is that its cloud-based scalability allows the system to handle large volumes of data, making it suitable for widespread deployment. The system leverages advanced Natural Language Processing (NLP) techniques to accurately identify fake news, while its real-time detection capabilities provide instantaneous results, enabling timely intervention and informed decision-making.

The proposed system's flexibility and scalability make it well-suited for expansion and seamless integration. A significant enhancement is multilingual support, enabling the system to process information in multiple languages and serve diverse global audiences. Designed to address the international and multilingual nature of fake news, the system can also improve claim verification accuracy by incorporating resources from trusted news and fact-checking organizations. Its modular architecture allows the integration of advanced AI models and real-time analytics tools, ensuring adaptability to emerging technologies. Together, these capabilities position the system as a powerful tool in combating disinformation

RESULTS AND DISCUSSION

To promote transparency and user understanding, the system incorporates advanced data visualization and reporting tools. Dashboards display metrics such as claim volume, accuracy, and the spread of flagged content, while interactive visualizations highlight misinformation trends, including topic- or region-specific false claims. Figure 6 presents ClaimBuster model performance for five sample claims, showing Precision, Recall, F1-Score, Accuracy, and False Positive Rate.

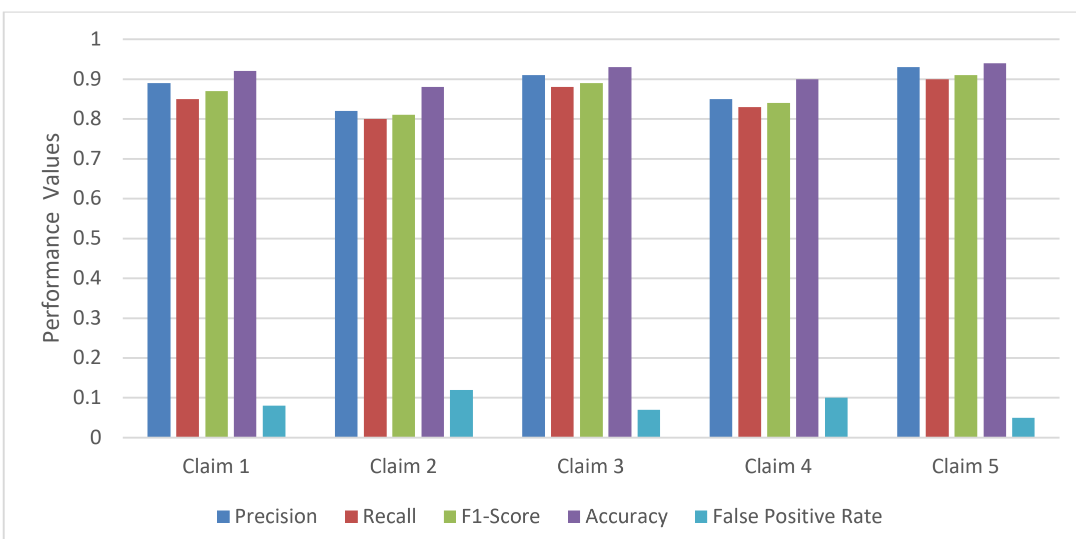

FIGURE 6. Model Performance Metrics

Precision indicates the proportion of correctly identified true claims, and Recall measures the model's ability to capture all relevant true claims. F1-Score combines Precision and Recall into a single metric, while Accuracy reflects the overall correct classification rate. The False Positive Rate measures the likelihood of misclassifying genuine claims as false. This comprehensive data provides insight into the model's strengths and weaknesses, with low False Positive Rates and high Precision and F1-Scores demonstrating its effectiveness at distinguishing true claims from false ones.

The scalability of the cloud platform allows ClaimBuster to handle increasing demands for real-time news analysis efficiently. Additionally, the system leverages advanced Natural Language Processing (NLP) techniques to ensure accurate claim detection, minimizing errors and reducing false positives. This combination of cloud scalability and sophisticated NLP enhances both the reliability and responsiveness of the system in combating misinformation. Gamification enhances user engagement and collaboration in combating disinformation by rewarding users who verify claims or provide contextual information. Badges, points, and leaderboards encourage friendly competition and sustained participation, motivating users to contribute actively

Leveraging cloud infrastructure significantly improves detection speed, enabling rapid identification of misinformation across large datasets. Cloud-based NLP algorithms enhance the accuracy and reliability of fake news detection, while adaptive resource management optimizes system performance and reduces operational costs. This integration of ClaimBuster with cloud computing ensures a scalable, efficient, and dependable solution for combating digital disinformation. The system offers tailored training courses to enhance its effectiveness for media professionals, educators, and journalists. These courses cover skills such as data visualization, claim verification, and the ability to recognize and reject false information. Figure 7 shows the usefulness of the ClaimBuster model's characteristics for identifying false news

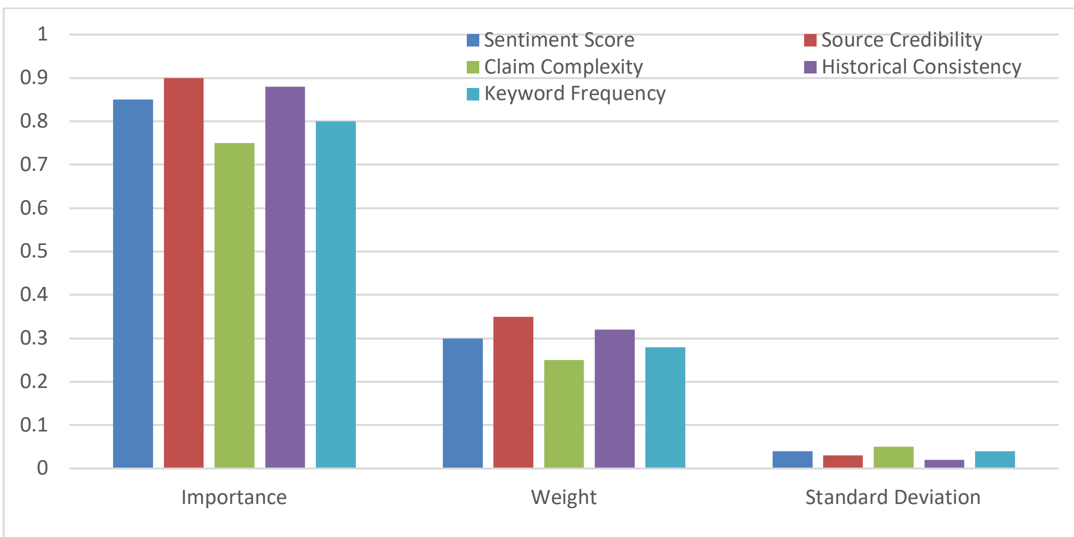

FIGURE 7. Feature Importance Scores

Features like Sentiment Score, Source Credibility, and Claim Complexity are evaluated using Importance Score, Weight, Standard Deviation, and Accuracy Impact. Source Credibility has the highest Importance Score of 0.90, highlighting its significant influence on the model's correctness. Weight indicates the proportion of each feature in the model, while Standard Deviation reflects the consistency of the feature across samples. Each attribute contributes to the model's predictive ability, collectively shaping its accuracy in detecting false claims. The system fosters partnerships between the public, private, and academic sectors to maximize efficiency. During public health emergencies or elections, this approach can help authorities trace and address misinformation. Collaboration with businesses, such as technology providers and content platforms, facilitates the integration of tools and processes. Academic cooperation supports research on misinformation and the advancement of fact-checking technologies. By working together, these groups enhance the system's capabilities and make it more responsive to emerging challenges. The fight against misinformation benefits from the combined expertise and resources of diverse organizations.

CONCLUSION

This study highlights the potential of combining ClaimBuster with cloud computing to address the growing challenge of fake news in the digital era. Cloud-based architecture ensures scalability, flexibility, and real-time responsiveness, enabling effective detection of misinformation across diverse platforms and contexts. Experimental results confirm the system's high accuracy and reliability, supported by adaptive machine learning models and continuous dataset updates. Ethical considerations, including data privacy, transparency, and bias mitigation, further strengthen the system's credibility and user trust. Future enhancements, such as multilingual support, sentiment analysis, and broader integration with social media platforms, will expand its applicability and effectiveness. By fostering collaboration across public, private, and academic sectors, the proposed solution provides a powerful and sustainable approach to safeguarding information integrity and mitigating the harmful effects of disinformation on society.

REFERENCES

[1]. S. Kotitsas, P. Kounoudis, E. Koutli, and H. Papageorgiou, 2024, "Leveraging fine-tuned large language models with LoRA for effective claim, claimer, and claim object detection," *18th Conference of the European Chapter of the Association for Computational Linguistics*, 1, pp. 2540-2554.

[2]. A. Shah, H. Shah, V. Bafna, C. Khandor, and S. Nair, 2024, "VERITAS-NLI: Validation and extraction of reliable information through automated scraping and natural language inference," *arXiv preprint arXiv:2410.09455*, vol. 147, pp. 1-15.

[3]. V. Setty, 2024, "Surprising efficacy of fine-tuned transformers for fact-checking over larger language models," *47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2842-2846.

[4]. V. Venktesh, A. Anand, A. Anand, and V. Setty, 2024, "QuanTemp: A real-world open-domain benchmark for fact-checking numerical claims," *47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 650-660.

[5]. A. Kan, C. Kan and Z. Nabulsi, 2024, "ViMGuard: A novel multi-modal system for video misinformation guarding," *arXiv preprint arXiv:2410.16592*, pp. 1-7.

[6]. A. Anand, A. Anand, and V. Setty, 2024, "NUMTEMP: A real-world benchmark to verify claims with statistical and temporal expressions," *arXiv preprint arXiv:2403.17169*, pp. 1-11.

[7]. J. R. Martinez-Rico, L. Araujo, and J. Martinez-Romo, 2024, "Building a framework for fake news detection in the health domain," *Plos One*, 19(7), Article. e0305362.

[8]. A. Seelam, A. Paul Choudhury, C. Liu, M. Goay, K. Bali, and A. Vashistha, 2024, "Fact-checks are for the top 0.1%: Examining reach, awareness, and relevance of fact-checking in rural India," *ACM on Human-Computer Interaction*, 8(1), pp. 1-34.

[9]. Y. Fu, S. Guo, J. Hoffswell, V. S. Bursztyn, R. Rossi, and J. Stasko, 2024, "The data says otherwise—towards automated fact-checking and communication of data claims," *37th Annual ACM Symposium on User Interface Software and Technology*, pp. 1-20.

[10]. I. Srba, O. Razuvayevskaya, J. A. Leite, R. Moro, I. B. Schlicht, F. M. García, S. B. Lottmann, D. Teyssou, V. Porcellini, and C. Scarton, 2024, "A survey on automatic credibility assessment of textual credibility signals in the era of large language models," *arXiv preprint arXiv:2410.21360*, pp. 1-66.

[11]. C.-M. Lai, and Y.-H. Guo, 2024, "Tracking of disinformation sources: Examining pages and URLs," IEEE Transactions on Computational Social Systems, 11(5), pp. 6242-6253.

[12]. V. Papanikou, P. Papadakos, T. Karamanidou, T. G. Stavropoulos, E. Pitoura, and P. Tsaparas, 2024, "Health misinformation in social networks: A survey of IT approaches," *arXiv preprint arXiv:2410.18670*, pp. 1-40.

[13]. L. Kavtaradze, 2024, "Challenges of automating fact-checking: A technographic case study," *Emerging Media*, 2(2), pp. 236-258.

[14]. U. Gadiraju, V. Setty, and S. Buijsman, 2024, "Claim check-worthiness in podcasts: Challenges and opportunities for human-AI collaboration to tackle misinformation," *Association for the Advancement of Artificial Intelligence*, pp. 1-4.

[15]. A. Halitaj, and A. Zubiaga, 2024, "Providing citations to support fact-checking: Contextualizing detection of sentences needing citation on small Wikipediaes," *Natural Language Processing Journal*, 8, Article. 100093.

[16]. R. Setty, and V. Setty, 2024, "QuestGen: Effectiveness of question generation methods for fact-checking applications," *33rd ACM International Conference on Information and Knowledge Management*, pp.

4036-4040.

[17]. I. Vykopal, M. Pikuliak, S. Ostermann, and M. Šimko, 2024, "Generative large language models in automated fact-checking: A survey," *arXiv preprint arXiv:2407.02351*, pp. 1-22.

[18]. S. Nenno, 2024, "Potentials and limitations of active learning: For the reduction of energy consumption during model training," *Weizenbaum Journal of the Digital Society*, 4(1), pp. 1-29.