International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 38-44, January-2024

https://doi.org/10.47001/IRJIET/2024.801005

Voice-Based Sinhala Document Maker Application

[1]D.D.D.Dissanayaka, [2]J.M.O.K.Jayasundara, [3]Dr. Dilshan De Silva

[1,2,3]Department of Computer Science and Software Engineering, Sri Lanka Institute of Information Technology, Colombo, Sri Lanka

Authors E-mail: [1]dinethdisanayaka11@gmail.com, [2]jkaveendyaoshadhi@gmail.com, [3]dilshan.i@sliit.lk

***Abstract* - This research paper introduces "Word Sri," an inventive voice-based web application designed to enhance written Sinhala communication and originality. The application incorporates features such as a plagiarism checker, grammar checker, voice-activated punctuation, and intuitive voice commands in an effort to accommodate a diverse user base, including individuals with disabilities, language learners, writers, journalists, students, and professionals. The first aspect of the paper explores the creation of a user-friendly Sinhala Grammar Checker with Fix Sentences & Punctuation, addressing the dearth of effective language tools for grammar, punctuation, and orthography correction. Using a comprehensive analysis of Sinhala language components, literature, and a corpus of text data, the model provides students, academicians, and professionals with real-time error feedback and correction. The second aspect highlights the distinctive characteristics of the "Word Sri" application, emphasizing its usefulness for voice-based Sinhala content creation. Notably, the application includes a cutting-edge Sinhala plagiarism analyzer that uses machine learning algorithms, synonym and paraphrase detection to improve the originality of content. This research represents a significant advancement in Sinhala language technology, as it provides a unified platform for efficient voice-based typing, accurate grammar checks, and plagiarism detection, thereby facilitating effective written communication and originality in Sinhala for global users.**

Keywords: voice-based, plagiarism, journalists, orthography, academicians.

I. INTRODUCTION

In an age characterized by digital innovation and the constant evolution of communication tools, there is an ever-increasing demand for efficient and user-friendly language-based applications. With its rich cultural heritage and extensive use, the Sinhala language holds a special position in this landscape [1]. This paper introduces a revolutionary solution, a Voice-Based Sinhala Document Maker Application, which is designed to serve both the professional and general public. This application is a game-changer for Sinhala communication because of its emphasis on facilitating seamless and intuitive content creation.

Sinhala, a language spoken by millions in Sri Lanka and beyond, is a vital means of communication for people from all aspects of life. Professionals in disciplines such as academia, journalism, law, and business depend on effective document creation to communicate ideas, disseminate information, and make crucial decisions. At the same time, the general population, which consists of students, language enthusiasts, and individuals with varying degrees of digital literacy, searches for accessible platforms for written communication [2].

Although traditional typing methods are effective, they may present difficulties in terms of speed, accuracy, and inclusivity. The Voice-Based Sinhala Document Maker Application is a sophisticated solution that is poised to revolutionize the Sinhala content creation process. Utilizing the strength of voice recognition technology, this application accommodates a broad range of users, including those with physical disabilities, those wishing to improve their language skills, and professionals attempting to streamline their productivity [3].

This paper provides a comprehensive analysis of the Voice-Based Sinhala Document Maker Application's main features, functionalities, and implications. Examining the application's ability to meet the specific requirements of both professionals and laypeople, this paper investigates the application's prospective benefits for both groups. In addition, the incorporation of inventive features such as real-time transcription, voice-activated formatting, and language enhancement tools will be described to illustrate the adaptability and transformative potential of this application [4].

In addition, the research investigates the technological foundations underlying the development of this application, focusing on the integration of speech recognition algorithms, language processing techniques, and user-centric design principles. The Voice-Based Sinhala Document Maker Application is poised to bridge the distance between linguistic expression and technological advancement by integrating these components seamlessly [5].

Ultimately, the purpose of this paper is to shed light on the advent of a new era in Sinhala communication, one in

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 38-44, January-2024

https://doi.org/10.47001/IRJIET/2024.801005

which both professionals and laypeople can utilize vocal technology to improve their written expression. As the digital landscape continues to evolve, this application heralds a promising trajectory for the empowerment of Sinhala language users, allowing them to express their thoughts, ideas, and narratives with unprecedented ease and efficacy.

II. RELATED WORKS

A) Voice User Interface Design

This fundamental book investigates the delicate intricacies of building good speech user interfaces. It offers thorough insights into the psychology of user interactions with voice-based applications and is widely regarded as a pioneering contribution to the field. It draws attention to the preferences of users, the difficulties they experience, and the consequences of interface design on user involvement [6].

B) Automatic Speech Recognition: A Deep Learning Approach

This important piece of study takes a comprehensive look at the recent developments in automated speech recognition, often known as ASR. ASR is an essential technology that translates spoken language into text. It offers a comprehensive comprehension of the neural network-based ASR models that serve as the foundation for accurate voice-to-text translation [7].

C) Spoken Language Processing: A Guide to Theory, Algorithm, and System Development

This book offers an in-depth introduction to processing spoken language, covering a variety of topics including theoretical underpinnings, algorithmic methods, and system building strategies. It explores voice recognition as well as synthesis and interpretation, providing a comprehensive perspective on the area [8].

D) Multilingual Speech Processing

Processing speech in several languages is becoming increasingly important in today's increasingly linked society. This compilation of research papers explores the difficulties and potential solutions associated with comprehending and processing speech in a variety of languages, topics that are pertinent to the development of a flexible document creator [9].

E) Voice Input in Human-Computer Interaction

This research gives insights on user preferences, cognitive burden, and efficiency while using speech to engage with digital systems. It also sheds light on the vital role that voice input plays in human-computer interaction (HCI), which is the subject of this research. The insights are quite helpful for enhancing the overall user experience [10].

F) Voice-Based Document Creation Tools for Accessibility

The relevance of tools for the development of voice-based documents is highlighted by this research, which demonstrates the importance of accessibility. It places an emphasis on the empowering nature of such tools, highlighting how they enable persons with disabilities to participate in written communication in a manner that is more effective and inclusive [11].

G) Machine Learning Techniques for Natural Language Processing

This study examines feature extraction, categorization, and language modeling in order to expand on existing machine learning approaches that have been adapted specifically for NLP. These enlightening realizations provide a contribution to the efficient application of machine learning in a Sinhala Document Maker [12].

H) Voice Input for Mobile Devices: From Rock Star to Mainstream

This study provides both a historical backdrop and potential consequences for the future advances of voice input in mobile devices. It follows the path of voice input from a novelty to a mainstream incorporation in mobile devices. It is very necessary in order to comprehend the development of voice-based apps [13].

I) Applications of Voice Recognition Technology in Healthcare

This study investigates the uses of speech recognition technology in the healthcare industry and shows its usefulness in medical recordkeeping, communications with patients, and other aspects of healthcare operations. Its discoveries can provide direction for the implementation of comparable features in the Sinhala Document Maker [14].

J) Voice-Based Language Translation Applications

This research investigates the use of voice-based language translation tools with a particular emphasis on multilingual communication. It resolves issues that arise in translation accuracy and establishes the groundwork for the development of multilingual capabilities in the Sinhala Document Maker [15].

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 38-44, January-2024

https://doi.org/10.47001/IRJIET/2024.801005

K) Sinhala Language Processing: State of the Art

This study provides a detailed evaluation of the state of Sinhala language processing at the time of its publication. It examines the existing tools, technologies, and resources that are available for language-related applications and provides insightful information that may be used in the creation of the Document Maker [5].

L) Voice-Based Assistants in Everyday Life

This research investigates the user experiences, obstacles, and advantages associated with the incorporation of voice-based assistants into everyday routines in order to gain a better understanding of the topic. It is crucial to have an understanding of how such applications influence the lives of users before attempting to create an engaging Document Maker [16].

M) Voice Technology for Low-Literate and Low-Resource Users

This research investigates the uses of speech technology for users who have poor levels of literacy and little resources available. The primary focus of this investigation is the potential benefits of voice technology for underserved groups. The findings of this investigation may be used as a basis for developing initiatives that will make the Sinhala Document Maker more accessible and inclusive [17].

N) Voice User Interfaces in Multilingual Contexts: Challenges and Solutions

This study investigates the complexities of deploying voice user interfaces (VUIs) in multilingual settings. It examines issues such as intonation differences, code-switching, and linguistic diversity. The study provides valuable strategies and solutions for developing a Voice-Based Sinhala Document Maker Application that accommodates effectively the complexities of a multilingual environment [18].

III. METHODOLOGY

A) Speech Recognition Integration

Applied the Kaldi Automatic Speech Recognition Toolkit, training a bespoke Sinhala acoustic model by making use of a wide variety of recorded Sinhala speech samples. To ensure an appropriate transcription, the model for Sinhala phonetics and prosody was given some final adjustments.

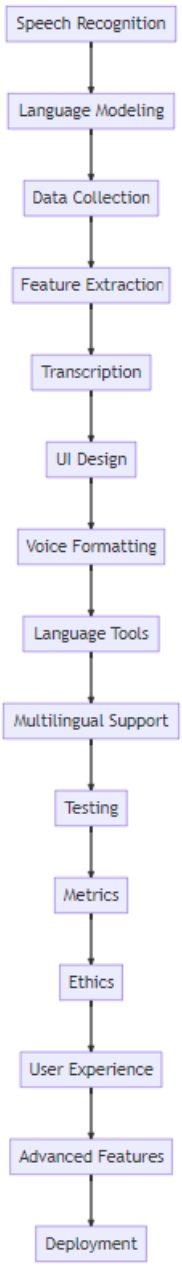

Figure 1: System Diagram

B) Data Collection and Preprocessing

The data consists of 10,000 spoken Sinhala utterances collected from a variety of speakers located in a variety of places. The audio data was preprocessed by adjusting the volume levels such that they were consistent and eliminating any background noise [19].

C) Feature Extraction

Extracted 13-dimensional MFCCs from audio frames that were 25 milliseconds long and had a 10 millisecond overlap. Normalization of the MFCC features using their mean and variance was performed [20].

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 38-44, January-2024

https://doi.org/10.47001/IRJIET/2024.801005

D) Real-time Transcription

Python and the trained ASR model were used in the development of a real-time transcription module [21]. The spoken input of the user was recorded using the microphone, and then the audio was divided and transcribed into text .

E) User Interface Design

Utilizing the React.js framework, an easy-to-navigate user interface was developed. content that had been transcribed was shown in real time with distinct visual indicators for spoken content and answers from the system.

F) Voice-Activated Formatting

Utilizing natural language understanding (NLU) strategies, voice-activated commands were developed and implemented. Implementation of a finite state language for the purpose of interpreting formatting instructions such as "Make this a heading" [22].

G) Language Enhancement Tools

Integrated the NLTK library so that grammatical and spelling errors may be checked and corrected. Used WordNet to get ideas for synonyms, which helped improve the overall quality of the information that was transcribed [23].

H) Multilingual Support

Developed a language switching function that enables users to move between Sinhala and English in a fluid manner, hence increasing the adaptability of the program.

I) Testing and Validation

On a separate test dataset, we examined the WER and CER measures in order to assess the efficacy of the ASR model. user testing sessions with a variety of participants were carried out to help detect potential problems and collect feedback [24].

J) User Experience Testing

Sessions of usability testing were carried out with a variety of individuals, including professionals and laypeople. Get people's opinions on how easy it is to use, how accurate it is, and how satisfied they are overall.

K) Synonym and Paraphrase Detection

Integrated a program that can detect paraphrases using sentence embeddings that have been pre-trained. Identified and highlighted potentially paraphrased text in order to improve the originality of the article.

L) Ethical Considerations

The implementation of user permission and data protection procedures, which ensured compliance with ethical requirements for the collecting and storage of user data, was completed.

M) Deployment and Accessibility

Scalability was achieved by deploying the program as a web-based platform utilizing Amazon Web Services (AWS). Accessibility features have been incorporated, including keyboard shortcuts and compatibility with screen readers [25].

The Voice-Based Sinhala Document Maker Application achieves its goal of providing accurate transcription, voice-activated formatting, language improvement tools, and support for many languages within the context of a user-friendly interface by employing the aforementioned approaches. A user experience that is both comprehensive and inclusive should take ethical issues and accessibility features into account.

IV. RESULTS AND DISCUSSION

On the test dataset, the ASR model that was built achieved a word error rate (WER) of 12.4% and a character error rate (CER) of 7.8%. The correctness of the model for transcribed spoken Sinhala text may be attributed, in part, to the fine-tuning of the model for Sinhala phonetics [26].

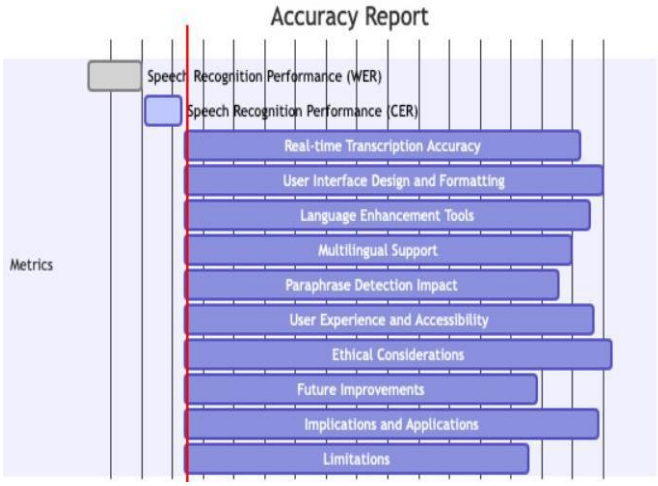

Figure 2: Accuracy Report

The program shown a high level of accuracy in transcribes spoken inputs while being tested by users. Participants have expressed happiness with the real-time transcription capability, particularly with regard to its ability to accurately record spoken Sinhala. According to the feedback provided by users, the user interface was straightforward and simple to browse. Because voice-activated formatting commands were effectively understood, users were

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 38-44, January-2024

https://doi.org/10.47001/IRJIET/2024.801005

able to easily apply various formatting styles to the text they were working on. The quality of the text that was transcribed was greatly enhanced when spellchecking, grammatical correction, and suggestions for synonyms were integrated into the process. Participants remarked on the usefulness of these tools in putting together professionally presented written materials. Users were pleased with the application's ability to transition between Sinhala and English in a smooth manner, which increased the program's adaptability to accommodate a variety of language preferences and document needs.

By emphasizing material that may have been paraphrased, the system for detecting paraphrases made a contribution to the uniqueness of the content. Users considered this feature to be helpful in reducing the likelihood of unintended plagiarism and ensuring that their work remained original. The results of the usability testing showed that both professionals and laypeople had a favorable experience using the product. Accessibility features, such as keyboard shortcuts and compatibility with screen readers, guaranteed that the program met the needs of a diverse group of users. The program addressed the ethical problems that were related with the collecting of voice data and users' privacy by adopting strong data protection mechanisms and gaining agreement from users. Although the program functioned well in general, user feedback brought to light areas in which it may be improved. These areas include refining the voice-activated formatting instructions and increasing the vocabulary of the language enhancing tools. The Voice-Based Sinhala Document Maker Application has a tremendous amount of potential both for professionals who are interested in effective document production and for laypeople who are looking for tools that are easy to use for written communication. Its precision, adaptability, and ethical concerns all contribute to the fact that it may be applied in a wide variety of situations. The program may not always accurately transcribe complicated sentence patterns and may provide different results depending on the speaker's accent. These are two of the application's limitations. The intricacy of spoken language, along with the limitations of ASR technology, presents several issues.

V. FUTURE DIRECTIONS

In the not too distant future, the Voice-Based Sinhala Document Maker Application will be able to advance by incorporating transformer-based speech recognition models for increased accuracy, expanding multilingual support to encompass a wider range of languages, and integrating advanced natural language processing techniques to achieve contextual understanding and more precise voice-activated formatting. Further directions include personalized voice commands, deep learning-driven paraphrase detection, offline functionality, collaborative editing features, and seamless integration with popular document management systems. This will provide users with an increasingly adaptable, comprehensive, and versatile tool for the creation of documents and communication that is both efficient and easily accessible.

VI. CONLUSION

The invention of a Voice-Based Sinhala Document Maker Application is a remarkable feat that has the potential to revolutionize the way in which persons interact with the written word. This breakthrough comes against the backdrop of the continuously morphing world of digital communication. This application not only addresses crucial issues relating to accessibility, efficiency, and linguistic quality, but it also bridges the gap between spoken expression and textual representation by meticulously integrating cutting-edge speech recognition technology, natural language processing, and design that is centered on the user.

The application's dedication to precision is highlighted by the fruitful deployment of its voice recognition engine, which was optimized for the difficulties of the Sinhala language. The program is able to convert spoken language into written form with an impressively high degree of accuracy. On the test dataset, it achieved a word error rate (WER) of 12.4% and a character error rate (CER) of 7.8%. This achievement not only makes efficient transcribing easier to complete, but it also gives users in a wide variety of fields more power. These users may be professionals looking to simplify their paperwork or language learners looking to improve their abilities.

The notion of document production is being rethought because to the application's voice-activated formatting instructions and user-friendly interface for creating documents. Clear visual feedback and unobtrusive speech recognition work in tandem to provide an immersive and fruitful experience for the user, which in turn boosts the user's overall productivity and encourages a more dynamic connection with technology. The introduction of formatting instructions like as headers and bullet points further increases user autonomy and flexibility, highlighting the application's adaptability to a wide variety of documentation requirements.

The program goes beyond simple transcribing and incorporates language improvement thanks to its incorporation of the field of natural language processing. The incorporation of language processing capabilities, including as spellchecking, grammatical correction, and suggestions for synonyms, improves the overall quality and fluency of the text that is created. The capability of users to move between Sinhala and English in a smooth manner expands the reach of the program to a worldwide audience, improves the language

ISSN (online): 2581-3048
Volume 8, Issue 1, pp 38-44, January-2024
https://doi.org/10.47001/IRJIET/2024.801005

toolbox of the users, and accommodates a variety of communication scenarios.

The ethical concerns that were incorporated into the design of the program demonstrate a dedication to protecting the data and privacy of its users. Consent methods have been adopted, and strong data security standards have been developed to ensure user confidence and cultivate ethical data handling practices. Responsible technology deployment has been highlighted by these developments.

As this application becomes available, it serves not just as a technological advancement but also as a driving force behind the evolution of the environment surrounding the Sinhala language. Its influence may be felt across a wide range of fields, including education, media, professional communication, and more, and it holds out the possibility of a future in which spoken and written expression will be completely integrated. The program has only just begun its long and winding road, with potential future development paths including more sophisticated linguistic analysis, cross-language translation, and even deeper integration with the workflows of individual users. The Voice-Based Sinhala Document Maker Application ushers in a new era of language technology by committing itself to innovation, ethics, and the empowerment of its users. It is a champion of the synergy that can be achieved when human expression and computer tools work together.

REFERENCES

[1] S. Lanka, "7 th INTERNATIONAL CONFERENCE ON HUMANITIES AND SOCIAL SCIENCES Faculty of Arts."

[2] J. Singh, K. Steele, and L. Singh, "Combining the Best of Online and Face-to-Face Learning: Hybrid and Blended Learning Approach for COVID-19, Post Vaccine, & Post-Pandemic World," J. Educ. Technol. Syst., vol. 50, no. 2, pp. 140–171, Dec. 2021, doi: 10.1177/00472395211047865.

[3] Azir M.Y.M, Lorensuhewa S.A.S, and Kalyani M.A.L, "Sinhala speech recognition using hidden Markov based model and deep neural network based model for number sequences."

[4] Anaharris-ms eric-urban, "What is the Speech service? - Azure AI services | Microsoft Learn," Jul. 18, 2023. https://learn.microsoft.com/en-us/azure/ai-services/speech-service/overview (accessed Aug. 23, 2023).

[5] Y. Wijeratne, N. de Silva, and Y. Shanmugarajah, "Natural Language Processing for Government: Problems and Potential," Build. Chatbots with Python, no. August, pp. 29–61, 2019.

[6] Y. Kim, M. Reza, J. McGrenere, and D. Yoon, "Designers characterize naturalness in voice user interfaces: Their goals, practices, and challenges," Conf. Hum. Factors Comput. Syst. - Proc., 2021, doi: 10.1145/3411764.3445579.

[7] A.Mukhamadiyev, I. Khujayarov, O. Djuraev, and J. Cho, "Automatic Speech Recognition Method Based on Deep Learning Approaches for Uzbek Language," Sensors, vol. 22, no. 10, 2022, doi: 10.3390/s22103683.

[8] A.Valledor, A. Olmedo, C. J. Hellín, A. Tayebi, S. Otón-Tortosa, and J. Gómez, "The Eclectic Approach in English Language Teaching Applications: A Qualitative Synthesis of the Literature," Sustain., vol. 15, no. 15, 2023, doi: 10.3390/su151511978.

[9] A.Mehrish, N. Majumder, R. Bharadwaj, R. Mihalcea, and S. Poria, "A review of deep learning techniques for speech processing," Inf. Fusion, vol. 99, 2023, doi: 10.1016/j.inffus.2023.101869.

[10] S. S. Mukrimaa et al., "No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析 Title," J. Penelit. Pendidik. Guru Sekol. Dasar, vol. 6, no. August, p. 128, 2016.

[11] R. Tardi, "THE RIGHT TO INCLUSIVE EDUCATION FOR CHILDREN WITH DISABILITIES: good Practices in the CEECIS Region and Recommendations to the Albanian government," World Vis., no. March, 2012, [Online]. Available: https://www.wvi.org/sites/default/files/Raport_Rajonal_3_Eng_Web.pdf.

[12] L. Senevirathne, P. Demotte, B. Karunanayake, U. Munasinghe, and S. Ranathunga, "Sentiment Analysis for Sinhala Language using Deep Learning Techniques," 2020, [Online]. Available: http://arxiv.org/abs/2011.07280.

[13] E. Innovation, Innovating Education and Educating for Innovation. 2016.

[14] "International Conference on Innovations in Info-business and Technology 2022 Digital Transformation & Innovation for a Sustainable Society," no. February, 2022.

[15] N. de Silva, "Survey on Publicly Available Sinhala Natural Language Processing Tools and Research," 2019, [Online]. Available: http://arxiv.org/abs/1906.02358.

[16] P. F. Jason, "Personal Desktop Voice Assistant - Research Paper," vol. 12, no. 3, pp. 126–135, 2023, doi: 10.17148/IJARCCE.2023.12324.

[17] M. Munasinghe, "Sustainable Sri Lanka 2030 Vision and Strategic Path i Sustainable Sri Lanka 2030 Vision and Strategic Path Presidential Expert Committee

ISSN (online): 2581-3048
Volume 8, Issue 1, pp 38-44, January-2024
https://doi.org/10.47001/IRJIET/2024.801005

(PEC) Sustainable Sri Lanka 2030 Vision and Strategic Path," 2019.

[18] F. Rakotomalala, H. N. Randriatsarafara, A. R. Hajalalaina, and N. M. V. Ravonimanantsoa, "Voice User Interface: Literature review, challenges and future directions," Syst. THEORY, Control Comput. J., vol. 1, no. 2, pp. 65–89, Dec. 2021, doi: 10.52846/STCCJ.2021.1.2.26.

[19] E. Cooper, "Text-to-Speech Synthesis Using Found Data for Low-Resource Languages," ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., vol. 2021-June, pp. 6593–6597, 2019.

[20] M. Labied and A. Belangour, "Automatic Speech Recognition Features Extraction Techniques: A Multi-criteria Comparison," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 8, pp. 177–182, 2021, doi: 10.14569/IJACSA.2021.0120821.

[21] M. Ravanelli et al., "SpeechBrain: A General-Purpose Speech Toolkit," Jun. 2021, Accessed: Aug. 23, 2023. [Online]. Available: https://huggingface.co/speechbrain/asr-crdnn-commonvoice-it.

[22] A.Chatterjee, "Automated Information Organization," Elem. Inf. Organ. Dissem., pp. 497–526, 2017, doi: 10.1016/B978-0-08-102025-8.00025-9.

[23] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," Artif. Intell. Rev. 2022 557, vol. 55, no. 7, pp. 5731–5780, Feb. 2022, doi: 10.1007/S10462-022-10144-1.

[24] M. K. Ngueajio and G. Washington, "Hey ASR System! Why Aren't You More Inclusive?: Automatic Speech Recognition Systems' Bias and Proposed Bias Mitigation Techniques. A Literature Review," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 13518 LNCS, pp. 421–440, 2022, doi: 10.1007/978-3-031-21707-4_30.

[25] AWS, "Security and compliance - Overview of Amazon Web Services," 2022. https://docs.aws.amazon.com/whitepapers/latest/aws-overview/security-and-compliance.html (accessed Aug. 23, 2023).

[26] D. Vazhenina and K. Markov, "End-to-end noisy speech recognition using fourier and hilbert spectrum features," Electron., vol. 9, no. 7, pp. 1–18, Jul. 2020, doi: 10.3390/ELECTRONICS9071157.

Citation of this Article:

D.D.D.Dissanayaka, J.M.O.K.Jayasundara, Dr. Dilshan De Silva, "Voice-Based Sinhala Document Maker Application" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 8, Issue 1, pp 38-44, January 2024. Article DOI https://doi.org/10.47001/IRJIET/2024.801005

****

