Jurnal Pendidikan Multimedia (EDSENCE)

Journal homepage: https://ejournal.upi.edu/index.php/Edsence

Transcription of Informatics Final Project Seminar Recordings via Speech-to-Text

Trisna Gelar[1], Aprianti Nanda Sari[2], Djoko Cahyo utomo Lieharyani[3], Fauza Naylassana[4], and Nisrina Wafa Zakiya Hamdani[5]

Politeknik Negeri Bandung, Bandung, Indonesia
Correspondence: E-mail: trisna.gelar@polban.ac.id

ABSTRACT

Bandung State Polytechnic has implemented project-oriented problem-based learning as its new educational approach. Project 6, or Final Project, is a course that communicates student learning results. Documenting seminar events is crucial as it provides valuable resources for students to analyze seminar outcomes and address any inquiries from teachers. However, not all students are inherently proactive in documenting or recording these activities. The task of transcribing learning outcomes becomes distinct when the emphasis is placed on students. This study aims to develop and evaluate a speech-to-text model utilizing DeepSpeech for transcribing seminar presentations related to final year projects, tackling the difficulties presented by spontaneous speech patterns and specialized technical terminology in software engineering. The model is trained and assessed utilizing Word Error Rate (WER) and Character Error Rate (CER) measures. The results of this study are the development of speech-to-text systems for educational purposes, especially within project-based student-centered learning. These resulting transcriptions could benefit both students and educators by offering a searchable and analyzable account of seminar presentations and improving feedback.

ARTICLE INFO

Article History:
Submitted/Received 16 Sept 2024
First Revised 23 Nov 2024
Accepted 01 Dec 2024
First Available online 17 Dec 2024
Publication Date 17 Dec 2024

Keyword:
Speech To Text,
Final Project,
DeepSpeech,
Seminar Recording.

1. INTRODUCTION

The Speech to Text algorithm is utilized to generate transcriptions in several domains including agriculture (Gelar & Nanda, 2022), law (Tumminia et al., 2021), tourism (Higgins & Ikeda, 2021), health (Latif et al., 2021), and education. In the education sector, STT technology possesses significant promise for improving teaching and learning experiences (Bakken et al., 2019; Millett, 2021). It can be employed for real-time lecture captioning, transcript generation for recorded lectures, language acquisition assistance, and accessibility for students with hearing disabilities (Zhao et al., 2018). This technology enables students with varied learning styles and requirements to access material more effectively and efficiently. Also, STT employed to offer quantitative feedback while learning to speak Japanese language (Kataoka et al., 2019). However, STT still employed in teacher-centered learning settings, wherein the focus is on the teacher instructing the material without any collaborative learning.

The Bandung State Polytechnic Computer and Informatics Engineering Department is now adopting Project-Oriented Problem Based Learning (Firdaus et al., 2024; Setiarini & Wulan, 2021), especially in Project-based courses. The Project Based Learning Lab is now utilized for various Project courses, including Project 1: Desktop Software Development, Project 2: Library Based Application Development, Project 6 Final Project, and Cloud Computing. The Project 6 Final Assignment course primarily emphasizes the communication of student learning outcomes using a presentation format as a learning tool. To successfully finish this course, students must deliver four presentations in the form of seminars. These seminars are specifically referred to as Seminar 1, Seminar 2, Pre-Final Project Defense, and Final Project Defense (Sari et al., 2023).

Project-Oriented Problem Based Learning approach emphasizes student-centered learning, where students actively participate in projects that simulate real-world problem-solving scenarios. The final project course, a core component of this approach, requires students to present their project outcomes in a seminar format. Formally the presentation format consists of four sessions, namely opening (introduction of title and personnel), presentation (content of the material to be presented), questions and answers (teacher or other group of students) and closing (conclusion and greetings).

However, capturing and preserving the rich information exchanged during these seminars can be challenging; not all students are inherently proactive in documenting or recording these activities. Manual notetaking can be incomplete and time-consuming, and relying on students to document the proceedings can be inconsistent. Therefore, there is a need for an automated and reliable method to generate accurate transcriptions of these seminar presentations. Consequently, it is necessary to utilize tools or technology that assist students in documenting seminar activities. Subsequently, the recorded content is converted into written form using a speech-to-text system. There are algorithmic challenges for getting good transcription, especially in presentation and question and answer sessions, such as the type of conversation (spontaneous or not spontaneous), who speaks during the presentation, and the content during the presentation such as technical software engineering terminology.

The objective of this research is to develop a method for implementing speech to text technology to generate transcriptions from recorded Seminar 1 activities in the Final Project 6 Project course. The Final Assignment presentation format comprises four sessions: opening, presentation, questions and answers, and closure. The opening session involves introducing

DOI: https://doi.org/10.17509/edsence.v6i2.74660
p-ISSN:2685-2489 | e-ISSN:2685-2535

the title and individuals. The presentation session focuses on delivering the material to be presented. The questions and answers session allows lecturers or students from other groups to ask questions. The closing session involves concluding the presentation and offering greetings. However, this study specifically concentrated on the Presentation session at Seminar 1 to gather information from the presentation.

2. METHODS

This research consists of three stages, namely Initiation, Implementation and Evaluation. These stages can be seen in the following image.

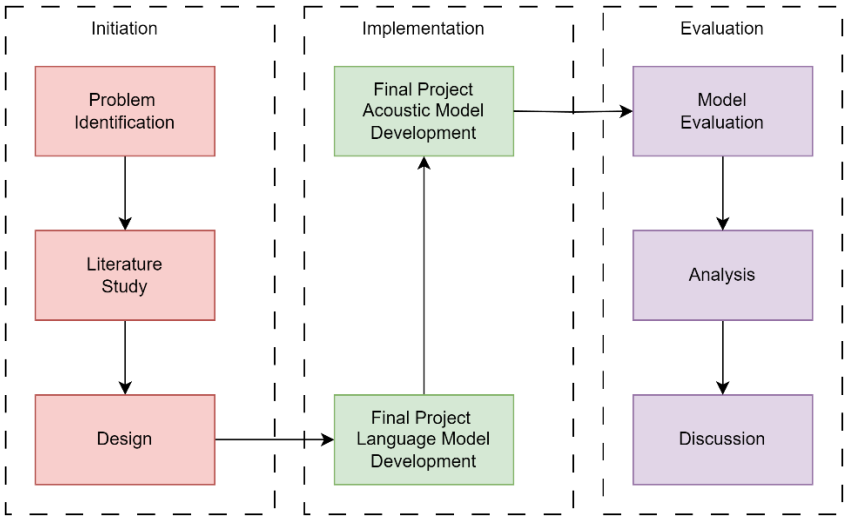

Figure 1. Research Methology

2.1. Initiation Step

The initial step comprises three components, specifically problem identification. Once the problem has been identified and its core cause discovered, the subsequent step involves doing a literature study. The objective is to acquire a foundational theory that is applicable to the problem at hand. In this literature study, researchers collected books, scientific articles, and other sources related to the following topics.

a) Final Project Proposal Seminar Problem Domain

The Final Project Text Document was sourced from the Bandung State Polytechnic DigiLib repository and the DIII Informatics Engineering Final Assignment Repository from the TA Coordinator to construct a language model. Every book is in PDF format and has a watermark, necessitating text processing. The processing is manually conducted by research assistant members, who are undergraduates. The excerpted topics from the TA book include Chapter I Introduction and Chapter II Literature Review, all of which have a Software Engineering focus. Text data cleaning includes automated processes such as case folding and punctuation mark removal. Manual

DOI: https://doi.org/10.17509/edsence.v6i2.74660
p-ISSN:2685-2489 | e-ISSN:2685-2535

tasks include converting numbers, symbols, and units into words, as well as deleting stop words.

Table 1. Example of a Final Assignment Corpus Sentence.

Sentence
Dalam menjalankan usahanya Bostani menerapkan konsep penjualan business to customer yaitu menjual produk secara langsung kepada pelanggan dengan sistem pemesanan pre-order.
Pre-Processed Sentence
dalam menjalankan usahanya bostani menerapkan konsep penjualan business to customer yaitu menjual produk secara langsung kepada pelanggan dengan sistem pemesanan pre order

b) DeepSpeech

Multiple related articles emphasize the utilization of libraries for Speech-To-Text, specifically the work of Trisna, et al. This study investigates the application of Mozilla Deep Speech for the Indonesian language in the agricultural arena, specifically focusing on agricultural YouTube videos that involve spontaneous conversations. This research consists of two models: the language model and the Indonesian audio model. The language model is trained using a collection of agricultural instructional books sourced from the ministry of agriculture repository (Gelar & Nanda, 2022).

Pablo Angel and Jeremy Hajek conducted a comparative analysis on the utilization of Speech to Text libraries, including both in-house options like Mozilla DeepSpeech and cloud-based solutions such as the AWS Transcribe API. The instance employed is a video recording specifically designed for the purpose of English language acquisition (Fernández & Hajek, 2020). These two articles will serve as a source of inspiration for the Speech to Text methodology.

c) Evaluating the Effectiveness of Speech Recognition Technology

Transcription results can be measured using two metrics: Character error rate and Word error rate. In addition, its quality is mostly determined by its readability. WER evaluates the accuracy of predicting word sequences that are recognized properly, whereas CER is determined by phoneme sequences. Both measures, namely the derivatives of the Levenshtein distance formula, are valuable for assessing enhancements in acoustic models (Zielonka et al., 2023).

Following the completion of the literature study, the subsequent phase involves analysis, encompassing several aspects. Examination of the components in the final project proposal seminar Analyzing the elements in video transcription using Speech to Text technology. The design implemented includes:

a) Recording equipment design and Standard Operating Procedures (SOP)

The primary goals of building tools and recording SOPs are to document the conversion of chosen TA Books for D3 2020-2023 Chapters 1-2 into text-to-speech format (totaling 15 books) and to capture audio from seminar participants during

DOI: https://doi.org/10.17509/edsence.v6i2.74660
p-ISSN:2685-2489 | e-ISSN:2685-2535

material presentations and lectures (including 3 students and 2 lecturers, including the TA Coordinator). The following graphic illustrates the necessary supporting equipment for one seminar session.

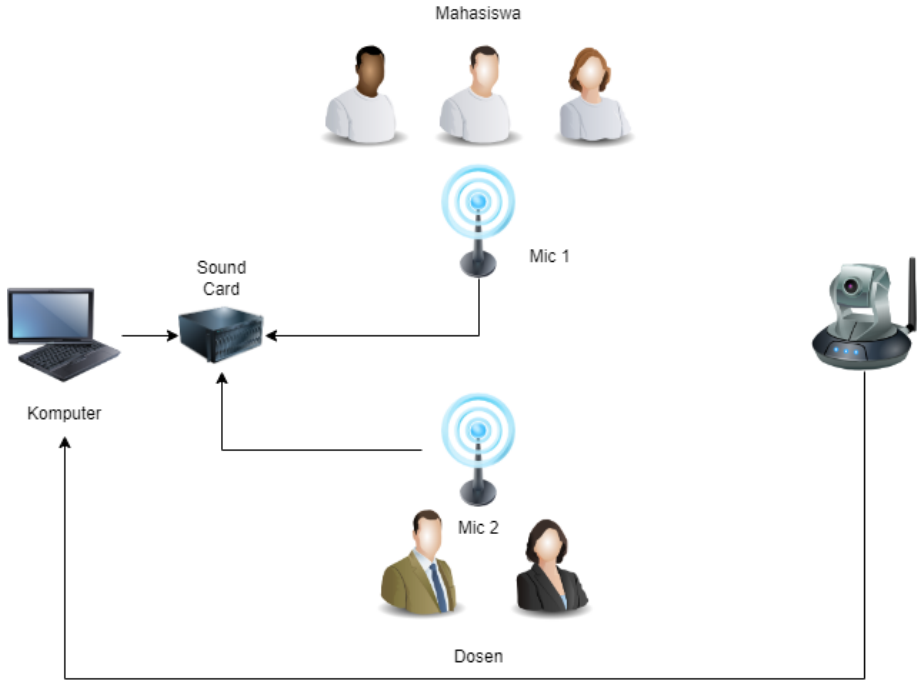

Figure 2. Design of a Device for Recording Seminars.

Featuring audio standards of 16 bits and the ability to record Full HD video. Two XLR type microphones with dynamic characteristics are connected to an external sound card, allowing for adjustable gain and volume settings to get consistent outcomes. The computer's recording application will be used to capture the audio. In addition, seminar activities will be captured on film as a point of reference. For instance, during question-and-answer sessions, you can record the speaker's identity and the corresponding timestamp. The SOP will standardize four crucial aspects: equipment usage processes, audio and video recording procedures, procedures for storing recorded data, and procedures for annotating data.

b) Materials and methodology design for the Language Corpus and Audio Corpus.

The acquisition of Language Corpus and Audio Corpus material involves the utilization of manual annotation. This annotation process consists of two distinct components:

1) Annotation of the Final Projects books chapters 1-2 for the years 2020-2023, intended for use as material for a language model.
2) The Final Projects Book Text Chapters 1-2 for 2020-2023 will be audio annotated manually by research assistants and researchers, intended for use as material for a language model.

DOI: https://doi.org/10.17509/edsence.v6i2.74660
p-ISSN:2685-2489 | e-ISSN:2685-2535

2.2. Implementation Step

During the implementation stage, there are two distinct components: language model development and acoustic model development. The first component, illustrated in Figure 3 consists of five steps. 1) the curation of Final Projects books for the period of 2020-2023. 2). Converting sentences from each page of an electronic book into text manually. 3). Conduct preprocessing sentences. 4) The texts are consolidated into a single document called the Final Project text corpus. 5) The text is transformed into a language model (LM) scorer using KenLM.

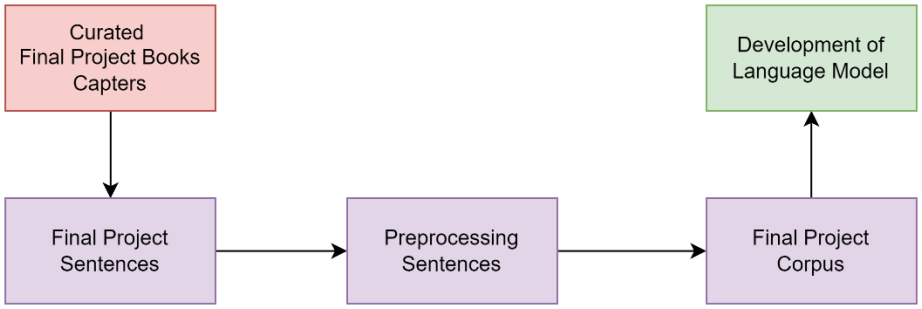

Figure 3. The process of developing a Final Project language model

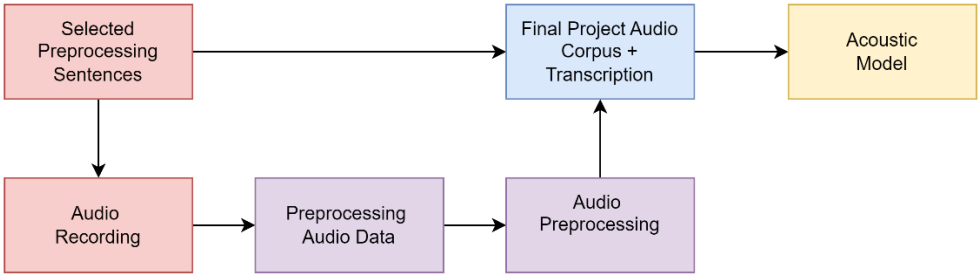

Figure 4. The process of developing a Final Project acoustic model

Figure 4 depicts the process of curating Final Project sentences audio into an acoustic model. There are four primary stages, specifically: 1) Selecting several Final Project books on catalog and do preprocessing sentences. 2) Audio Recording by Research Assistant 2 Female and 2 Male. 3) Preprocessing audio data such as increasing the voice volume, decreasing, or increasing tempo and remove long silent audio. 4) Once the audio corpus and transcriptions have been gathered, the file name, size, and transcription text details are stored in a CSV file. 5) The audio corpus and transcriptions will be used to construct acoustic models.

2.3. Evaluation Step

During the Evaluation step, the performance of language and acoustic modeling will be assessed based on WER and CER. The evaluation results will be reviewed to determine the optimal mix of components or identify any exceptional findings from the modeling results. Additionally, during the evaluation process, audio data augmentation was implemented through modifications in volume, pitch, and tempo to introduce changes to the final project's audio data. Subsequently, each finding undergoes a process of discussion.

DOI: https://doi.org/10.17509/edsence.v6i2.74660
p-ISSN:2685-2489 | e-ISSN:2685-2535

Table 2. Design Experimentation with three epoch combinations (10, 15, and 25)

	Original	Augment 01 (Volume)	Augment 02 (Pitch)	Augment 03 (Tempo)	Augment All
wer					
cer					
loss					
train time					
best wer					
worst wer					

3. RESULTS AND DISCUSSION

3.1. Corpus of Final Project Language

The Final Projects Language Corpus was constructed by extracting text from Chapter 1 and Chapter 2. Table 2 displays the status of book curation. To introduce different forms of Indonesian, the Final Project Language Corpus will be merged with the TED Multilingual Corpus dataset (https://github.com/ajinkyakulkarni14/TED-Multilingual-Parallel-Corpus) specifically for Indonesian.

Table 3. Results of the curation of the Final Project textbook

No	KoTA	Year	Chapter 1	Chapter 2
1	KoTA101, KoTA102, KoTA103, KoTA104, KoTA105, KoTA107, KoTA108, KoTA109, KoTA201, KoTA202, KoTA203, KoTA204, KoTA205, KoTA206, KoTA207, KoTA208, KoTA209, KoTA210, KoTA211	2021	Finished	Finished
2	KoTA101, KoTA102, KoTA103, KoTA104, KoTA105, KoTA106, KoTA107, KoTA109, KoTA110, KoTA111 KoTA202, KoTA203, KoTA204, KoTA205, KoTA206, KoTA207, KoTA208, KoTA209, KoTA210	2022	Finished	Finished
3	KoTA101, KoTA102, KoTA103, KoTA104, KoTA105, KoTA107, KoTA108, KoTA201, KoTA202, KoTA205, KoTA206, KoTA207, KoTA208, KoTA209, KoTA210	2023	Finished	Finished

3.2. The TA Acoustic Corpus is a collection of text documents used in the final project.

The Final Project acoustic corpus was created by four students (two females and two males) who recorded sounds from words in Chapter 1 or Chapter 2 of the TA book for the

DOI: https://doi.org/10.17509/edsence.v6i2.74660
p-ISSN:2685-2489 | e-ISSN:2685-2535

courses KoTA 101 and KoTA 107 in 2021, KoTA202 in 2021, KoTA 111 in 2022, and City 103 in 2023. In addition to the Final Project Acoustic Corpus, which specifically targets the Software Engineering field, this study also utilized the Indonesian language audio corpus from Common Voice version 18.0(https://commonvoice.mozilla.org/en/datasets), last updated on June 19, 2024. The corpus consists of a total of 65 hours of recorded material, 32 validated hours, 570 number of voices.

Table 4. Results of the curation of the Final Project textbook

No	Initial	Document	Chapter	Total Acoustic File
1	Female (01 & 02)	KoTA101 2021	1	147
			2	195
		KoTA107 2021	1	118
			2	175
		KoTA111 2022	1	54
			2	266
		KoTA103 2023	1	81
			2	171
2	Male 01	KoTA101 2021	1	145
			2	196
		KoTA107 2021	1	133
		KoTA202 2021	1	87
			2	57
3	Male 02	KoTA101 2021	1	145
			2	195
		KoTA107 2021	1	133
		KoTA202 2021	1	87

3.3. Assessment of the Speech to Text Model

Table 5. Model Evaluation with Epoch 10 without LM Scorer

	Original	Augment 01 (Volume)	Augment 02 (Pitch)	Augment 03 (Tempo)	Augment All
wer	1.000	1.0	1.0	1.00	1.00
cer	0.593	0.603	0.624	0.584	0.615
loss	112.929	114.952	115.242	114.074	114.676
train time	1:21:41	1:21:42	1:21:42	1:21:41	1:21:44
best wer	0.500	0.500	0.666	0.666	0.666
worst wer	3.000	4.00	4.00	3.50	3.500

Table 6. Model Evaluation with Epoch 10 with LM Scorer

	Original	Augment 01 (Volume)	Augment 02 (Pitch)	Augment 03 (Tempo)	Augment All
wer	0.928	0.990	0.996	0.912	0.993
cer	0.687	0.854	0.869	0.564	0.871

DOI: https://doi.org/10.17509/edsence.v6i2.74660
p-ISSN:2685-2489 | e-ISSN:2685-2535

loss	112.929	114.952	115.242	114.074	114.676
best wer	0.00	0.250	0.666	0.000	0.666
worst wer	3.50	1.0	1.0	6.500	1.0

Table 7. Model Evaluation with Epoch 15 without LM Scorer

	Original	Augment 01 (Volume)	Augment 02 (Pitch)	Augment 03 (Tempo)	Augment All
wer	1.0	1.0	1.00	1.0	1.00
cer	0.595	0.611	0.601	0.593	0.611
loss	114.305	114.778	113.07	114.621	114.77
train time	2:02:34	2:02:32	2:02:33	2:02:30	2:02:32
best wer	0.500	0.600	0.714	0.666	0.500
worst wer	4.0	4.000	3.000	5.500	3.00

Table 8. Model Evaluation with Epoch 15 with LM Scorer

	Original	Augment 01 (Volume)	Augment 02 (Pitch)	Augment 03 (Tempo)	Augment All
wer	0.976	1.00	0.996	0.993	1.00
cer	0.783	0.589	0.855	0.874	0.589
loss	114.305	114.778	113.074	114.621	114.778
best wer	0.500	0.333	0.500	0.667	0.00
worst wer	2.333	1.0	1.0	1.0	9.0

Table 9. Model Evaluation with Epoch 25 without LM Scorer

	Original	Augment 01 (Volume)	Augment 02 (Pitch)	Augment 03 (Tempo)	Augment All
wer	1.0	1.0	1.0	1.0	1.0
cer	0.566	0.577	0.591	0.604	0.585
loss	111.885	114.101	113.697	114.080	112.749
train time	3:24:09	3:24:11	3:24:13	3:24:12	3:24:19
best wer	0.500	0.500	0.600	0.500	0.666
worst wer	3.500	4.00	4.0	3.666	4.00

Table 10. Model Evaluation with Epoch 25 with LM Scorer

	Original	Augment 01 (Volume)	Augment 02 (Pitch)	Augment 03 (Tempo)	Augment All
wer	0.993	0.847	0.988	0.982	0.994
cer	0.863	0.579	0.820	0.781	0.875
loss	111.885	114.101	112.749	114.080	112.749
best wer	0.25	0.00	0.20	0.250	0.500
worst wer	1.0	5.00	1.0	2.33	1.0

DOI: https://doi.org/10.17509/edsence.v6i2.74660
p-ISSN:2685-2489 | e-ISSN:2685-2535

The WER evaluation values for every scenario from tables 5 to 9 exhibit minimal variation, with an average value approaching 1.0. Only Table 10 indicates that the WER value for every single case, both original and enhanced, is below 1.0. In contrast to WER, the CER value varies for every scenario from ± 0.5 to ± 0.88, although the value is better than WER, this value only shows pronunciation accuracy. Simultaneously, WER precisely predicts significant words and sentences.

Another parameter to consider is the best WER in the experimental scenario, where the Speech to Text system's Language Model Scorer can accurately predict the correct sentence with a WER value of 0.0. Table 11 exemplifies a statement with a WER of 0.0 across all scenarios. The combination of an epoch parameter set to 25 and volume augmentation yields the optimal WER within the Informatics Final Project domain.

Table 11. Examples of Best WER Sentences for each scenario

Scenario with LM Scorer	Sentences
Epoch 10 Original	semua makanan sudah tidak ada
	saya sudah membaca buku
	saya sudah membaca buku
	ini pertanyaan yang bagus
	aku seorang esperantis"
Epoch 10 Tempo	antara lain pemantauan antrean pasien
	jawaban pertanyaan ini ada tiga buah
	aku akan pergi ke kedai kopi
	yang dapat menangani pendaftaran pasien
	apa tidak ada yang lebih baik
Epoch 15 All	lepaskan tanganmu dari saya
	sekolah mulai senin depan
	konsultasi pasien
	aku menipumu
	pemberian obat
Epoch 25 Volume	sistem informasi pendataan anggota poliklinik upn veteran jawa timur
	yang datang ke poliklinik secara bersamaan
	yang telah diuraikan pada sub bab sebelumnya
	dia adalah seorang yahudi konservatif
	saya berbelanja di toserba dan supermarket

4. CONCLUSION

The research identifies four primary findings about the processes of data curation, data processing, model development, and evaluation. The initial step involves curating textual data from a compilation of Final Assignment books in PDF format obtained from the Final Assignment repository. Furthermore, a group of four students collected audio data by individually recording words from the chosen text, which will be utilized as an acoustic model. The acoustic data has been processed and formatted for training, validation, and testing the

DOI: https://doi.org/10.17509/edsence.v6i2.74660
p-ISSN:2685-2489 | e-ISSN:2685-2535

acoustic model. Similarly, the sentence data in the Final Project book has been converted into clean text and processed for developing the language model. The acoustic model was constructed using the Third, Final Project audio data and the Indonesian Common Voice dataset. A total of 15 training scenarios were executed, considering the utilization of data, augmentation type, and number of epochs. The language data for the final assignment has been converted into a language model score, which is utilized during evaluation to provide significance to the predicted words. Furthermore, the evaluation findings indicate that the model with the greatest epoch in each dataset does not consistently yield the optimal either WER or CER value. The acoustic model system with an LM scorer can yield WER values approaching 0.0 for several sentences, particularly the combo model of epoch 25 and volume augmentation, which accurately predicts sentences in the Final Project informatics domain.

The developed speech-to-text model possesses numerous practical applications in educational environments. It could be used to generate accurate transcripts of seminar presentations, lectures, and other academic debates. These transcripts can function as essential tools for students, allowing them to reread and understand the information at their own pace. Moreover, educators can employ the transcripts to evaluate student comprehension and deliver comments.

Despite its potential, the model possesses certain limitations. Factors such as background noise and speaker accents may influence the accuracy of the model. Moreover, incorporating recordings of discussion sections may augment the diversity of discourse as students and lecturers participate in discussions about final projects may generated cocktail party problems (Agrawal et al., 2023).

Future research may focus on enhancing the model's effectiveness by incorporating domain-specific language models and improving the acoustic model's robustness to noise. Integrating the suggested speech-to-text technology with current educational platforms and resources could provide a unified and advantageous tool for students and educators. This interface may facilitate automated captioning for seminars, the generation of searchable transcripts for recorded seminars, and personalized feedback systems.

ACKNOWLEDGMENT

The Research Program is funded by DIPA funds from the Bandung State Polytechnic, as stated in the letter of agreement PJBL Labs research aid with the number B/1.26/PL1.R7/PG.00.03/2024.

AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

DOI: https://doi.org/10.17509/edsence.v6i2.74660
p-ISSN:2685-2489 | e-ISSN:2685-2535

REFERENCES

Agrawal, J., Gupta, M., & Garg, H. (2023). A review on speech separation in cocktail party environment: challenges and approaches. *Multimedia Tools and Applications*, *82*(20), 31035–31067. https://doi.org/10.1007/s11042-023-14649-x

Bakken, J. P., Uskov, V. L., Rayala, N., Syamala, J., Shah, A., Aluri, L., & Sharma, K. (2019). *The Quality of Text-to-Voice and Voice-to-Text Software Systems for Smart Universities: Perceptions of College Students with Disabilities* (pp. 51–66). https://doi.org/10.1007/978-3-319-92363-5_5

Fernández, P. Á. Á., & Hajek, J. R. (2020). A Case Study in Comparative Speech-to-Text Libraries for Use in Transcript Generation for Online Education Recordings. *SIGITE 2020 - Proceedings of the 21st Annual Conference on Information Technology Education*, 223–228. https://doi.org/10.1145/3368308.3415380

Firdaus, L. H., Wulan, S. R., & Maspupah, A. (2024). A CASE environment for Project-based Course to learn a sustainable software development. *E3S Web of Conferences*, *479*, 07026. https://doi.org/10.1051/e3sconf/202447907026

Gelar, T., & Nanda, A. (2022). Exploration of Spontaneous Speech Corpus Development in Urban Agriculture Instructional Videos. *Journal of Software Engineering, Information and Communication Technology (SEICT)*, *3*(1), 1–14. https://doi.org/10.17509/seict.v3i1.44548

Higgins, C., & Ikeda, M. (2021). The materialization of language in tourism networks. *Applied Linguistics Review*, *12*(1), 123–152. https://doi.org/10.1515/applirev-2019-0100

Kataoka, Y., Thamrin, A. H., Murai, J., & Kataoka, K. (2019). Employing automatic speech recognition for quantitative oral corrective feedback in Japanese second or foreign language education. *ACM International Conference Proceeding Series*, 52–58. https://doi.org/10.1145/3369255.3369285

Latif, S., Qadir, J., Qayyum, A., Usama, M., & Younis, S. (2021). Speech Technology for Healthcare: Opportunities, Challenges, and State of the Art. *IEEE Reviews in Biomedical Engineering*, *14*, 342–356. https://doi.org/10.1109/RBME.2020.3006860

Millett, P. (2021). Accuracy of Speech-to-Text Captioning for Students Who are Deaf or Hard of Hearing. *Journal of Educational, Pediatric & (Re) Habilitative Audiology* , *25*, 1–13. https://www.edaud.org/journal/2021/1-article-21.pdf

Sari, A. N., Pribadi, D. S., Gelar, T., Rahmawati, A., Azzahra, N., & Oktoharitsa, H. (2023). Implementasi Digital Signature pada Laporan Tugas Akhir. *Jurnal Informatika Polinema*, *10*(1). https://doi.org/10.33795/jip.v10i1.1323

Setiarini, S. D., & Wulan, S. R. (2021). Analysis Software Engineering Team's Soft Skills Learning using Online Learning Platform with Project-Oriented Problem-Based Learning (POPBL). *Inform : Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, *6*(2), 81–86. https://doi.org/10.25139/inform.v6i2.3986

Tumminia, J., Kuznecov, A., Tsilerides, S., Weinstein, I., McFee, B., Picheny, M., & Kaufman, A. R. (2021). *Diarization of Legal Proceedings. Identifying and Transcribing Judicial Speech from Recorded Court Audio*. 5–9. http://arxiv.org/abs/2104.01304

Zhao, M., Zheng, W., Ye, Y., & Wu, M. (2018). *Research on Educational Video Retrieval Method Based on Audio Transcription Technology*. *149*(Mecae), 384–388. https://doi.org/10.2991/mecae-18.2018.75

Zielonka, M., Krasiński, W., Nowak, J., Rośleń, P., Stopiński, J., Żak, M., Górski, F., & Czyżewski, A. (2023). A survey of automatic speech recognition deep models performance for Polish medical terms. *2023 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 19–24. https://doi.org/10.23919/SPA59660.2023.10274442

DOI: https://doi.org/10.17509/edsence.v6i2.74660
p-ISSN:2685-2489 | e-ISSN:2685-2535

DOI: https://doi.org/10.17509/edsence.v6i2.74660
p-ISSN:2685-2489 | e-ISSN:2685-2535