JIKO (Jurnal Informatika dan Komputer) Accredited KEMENDIKTI SAINSTEK, No.0173/C3/DT.05.00/2025
Vol. 8, No. 2, August 2025, pp. 103-112 p-ISSN: 2614-8897
DOI: 10.33387/jiko.v8i2.10019 e-ISSN: 2656-1948

AUTOMATIC ABSTRACTIVE SUMMARIZATION OF CURRICULUM VITAE USING S-BERT AND T5

Reza Fahlevi Herdiyanto[1*], Dian Sa'adillah Maylawati[2], Nur Lukman[3]

[1,2,3]Program Studi Teknik Informatika, Universitas Islam Negeri Sunan Gunung Djati
Email: [*1]1217050120@student.uinsgd.ac.id, [2]diansm@uinsgd.ac.id, [3]n.lukman@uinsgd.ac.id

(Received: 04 June 2025, Revised: 20 June 2025, Accepted: 26 June 2025)

Abstract

The rapid advancement of technological disruption has catalyzed significant innovations in human resource management, particularly through the widespread adoption of automated applicant screening systems such as Applicant Tracking Systems (ATS). However, these systems often fail to identify potential candidates due to poorly formatted Curriculum Vitae (CV) or missing important keywords, resulting in many applicants being eliminated in the early stages of selection. This research aims to develop an automatic CV summarization system by utilizing Natural Language Processing (NLP) technology. This research uses a combination of Sentence-BERT (SBERT) algorithm for information extraction and Text-to-Text Transfer Transformer (T5) for text generation. The K-Fold Cross Validation method with k = 3 was used in the model performance evaluation, in accordance with the limited computing resources. Experimental results show that the SBERT model is able to extract important information with high accuracy (F1-score of 0.8866), while the T5 model is able to generate informative summaries with a ROUGE-1 score of 0.8680. The combination of SBERT in producing important information extraction from CV and T5 that produces an abstractive summary shows good results with ROUGE-1 scores of 0.5497, ROUGE-2 of 0.3537, and ROUGE-L of 0.4334. This system is able to produce CV summaries that make it easier for companies to select job applicants according to the criteria and increase the chances of applicants to pass the initial selection stage.

Keywords: *Applicant Tracking System, Curriculum Vitae, Text Summarization, SBERT, T5.*

**Corresponding Author: Reza Fahlevi Herdiyanto*

1. INTRODUCTION

Technological developments in the field of job recruitment have encouraged the integration of technology into the recruitment process. The process no longer involves humans in some of its stages. These stages include screening candidates using Curriculum Vitae (CV), automated interviews, and interactive chatbots [1][2]. The first stage that job applicants will go through is screening candidates using CVs.

Curriculum Vitae has a very important role in the process of applying for a job. Therefore, its creation should not be done carelessly, but must be adjusted to the keywords contained in the job vacancy. At this selection stage, the tool that is usually used is the Applicant Tracking System (ATS). ATS is a tool used to filter applicant CVs by matching certain keywords in the document [3].

The main role of an ATS is to manage, automate, and improve the recruitment process, making it easier to find, monitor, and screen job applicants [4][5]. As such, it can speed up the selection process and improve efficiency in managing job applicants. However, there are some major challenges in using ATS systems. Firstly, limitations in keyword matching can lead to bias and overlook candidates who are qualified, but did not use the right keywords [6]. The system also struggles to handle various resume formats, and faces obstacles in terms of integration, user adoption, and implementation and maintenance costs [4][6]. In addition, ATS tends to prioritize the needs of recruiters, leaving applicants unclear as to whether their resumes meet the criteria or not [7].

For this reason, the presentation of important information in the CV should be prioritized such as key words that will be extracted by the ATS system. These systems usually extract key points such as skills, experience, and education (degree) [8][9]. Customizing the CV with these points allows the document to be more easily read and processed by the

ATS system. Therefore, understanding how an ATS works is important for applicants to develop an effective job search strategy.

In an effort to increase candidates' chances of passing to the next selection stage, the development of a Curriculum Vitae (CV) summarization system can help optimize documents to better suit the needs of the ATS system. The use of Sentence-BERT (SBERT) and Text-to-Text Transfer Transformer (T5) models is an important component in this research. SBERT is able to accelerate the process of grouping and searching for information, making it more efficient in data extraction [10]. Meanwhile, T5 has the ability to generate summaries by retaining information from the input [11]. The combination of the two models is expected to be an effective step in the development of a CV summarization system.

Text summarization and resume filtering have been the focus of a number of previous studies. One of them showed that utilizing SBERT in the resume screening process can improve accuracy and efficiency. The study recorded a filtering time of 0.233 seconds per resume and up to 90% accuracy in identifying relevant candidates [12]. Document summarization using the T5 model is one of the effective approaches. In a previous study, the use of the fine-tuned T5 model on XSum and Gigaword news article datasets of 304,405 data for abstract summarization resulted in the best evaluation scores on ROUGE-1, ROUGE-2, and ROUGE-LSUM metrics of 43.02; 37.43; and 37.49, respectively [13].

Meanwhile, another study used the Bidirectional Encoder Representations from Transformers (BERT) model for extractive-abstractive summarization with the CNN/DailyMail dataset and obtained ROUGE-1 values of 31.82, ROUGE-2 of 10.81, and ROUGE-L of 27.51 [14]. In addition, another approach has also combined Local Outlier Factor (LOF), Sentence-BERT, and T5 in extractive summarization using fact-check report datasets from PolitiFact, Demagog, and SumeCzech news dataset of 77,866 data, and resulted in ROUGE-1 score of 40.76, ROUGE-2 of 22.00, and ROUGE-L of 38.36 [15].

Although SBERT and T5 have been used in previous studies for fact-check report summarization [15]. their application is still limited to claims verification and has not been directed towards job application optimization. Most studies also emphasize the quality of the summary results, without considering its readability by automated selection systems such as ATS. This research aims to develop an SBERT and T5-based CV summarization system that can summarize CVs and be adapted to the ATS evaluation mechanism, thereby increasing the chances of applicants passing the initial stage of selection.

2. RESEARCH METHOD

This research adapts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework as a research activity. CRISP-DM is a methodology used for structured project management. This methodology consists of six stages, starting from business understanding, followed by data understanding, data preparation, modeling, evaluation, and ending with the deployment stage [16]. The stages involved in this process are outlined in Figure 1.

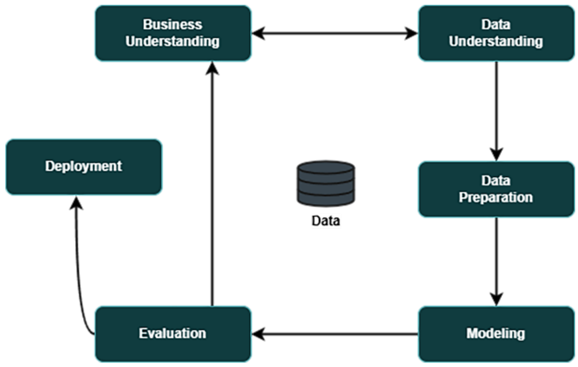

Figure 1. CRISP-DM Methodology [17]

2.1. Business Understanding

This stage is called problem understanding because it focuses on exploring the business problems to be solved. The main problem in this research is that many job applicants fail at the Curriculum Vitae (CV) assessment stage because the CV content does not match the job vacancy criteria, especially in companies that use the ATS (Applicant Tracking System) system. These systems automatically assess and extract important information from CV, such as skills, education, work experience, and certifications. Therefore, the aim of this research is to develop a CV summarization system that can highlight these points, so that applicants can evaluate the suitability of their CV to the job vacancy criteria and increase their chances of passing the selection process.

2.2. Data Understanding

Once the problem and objectives are understood, the next step is to collect and explore the data, including structure identification, quality evaluation, and initial pattern discovery to ensure data feasibility. This study used 1,116 PDF CV files obtained from Hugging Face and Kaggle [18][19], representing various professions such as accountant, advocate, architect, data science, HR, and others. The data was collected from March 1-18, 2025, with a screening process to ensure formatting and completeness of information such as personal data, skills, education, experience, and certifications. Although there were some incomplete CVs, overall the data obtained provides a clear picture of the diversity of CV structure, content, and characteristics from various professional fields, which is an important basis for the next stage of analysis.

2.3. Data Preparation

This stage is very important because it includes various data processing steps, from cleaning, transforming, to dividing the data into training and validation sets. The goal is for the model to learn from quality data. Because this research uses two models, the dataset is prepared in two different datasets according to the needs of each model. The stages performed in the data preprocessing process are shown in Figure 2.

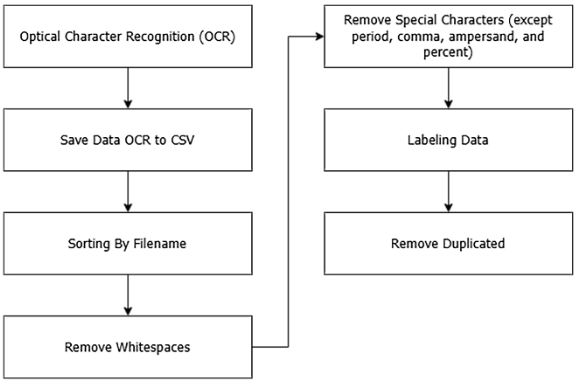

Figure 2. Data Preparation Stages

In Figure 2, there are seven main steps performed in the data preprocessing process. The process begins with text extraction from PDF files using Optical Character Recognition (OCR) techniques operated through the pytesseract library. The text extraction results are then saved in CSV format, which facilitates further processing and analysis. The extracted data is then cleaned by removing excess spaces, special characters, and irrelevant elements. After that, the data is sorted and converted into JSON format. This format was chosen because the JSON structure is more efficient for the labeling process with multiple categories in one entry, consisting of personal_information, skills, education, experience, and certification.

After labeling, duplication removal reduced the dataset from 1,116 to 1,007 entries. This process was performed in two stages, with the data divided into two parts. Duplications were removed in each part to avoid repetitive information that could affect model training. After merging the two parts, a final check was performed to ensure optimal data quality before being used in model training.

The first dataset was used to train the SBERT model as a span classifier, while the second dataset was intended for the T5 model in the text generation task. The format of the second dataset is organized with two keys, *extracted*_information and *expected_generation.* Once both datasets were ready, the data was divided into training and validation data, with an experimental 70:30 split, different from the commonly used 80:20 split [15].

2.4. Modeling

In the modeling stage, the method is selected according to the purpose of the analysis and the characteristics of the data, with the model trained using the prepared dataset. The focus of model development is two key tasks, which are information extraction from CVs and generative text. For information extraction, a pre-trained Sentence-BERT (SBERT) model is used to convert the OCR text into points according to the label. As for the generative text, the Text-to-Text Transfer Transformer (T5) model is used to generate text based on the extracted information. The following is an explanation of the model architecture that will be used in this research:

1) Sentence-BERT (SBERT)

The SBERT model "all-mpnet-base-v2" was chosen because it shows good performance and the selection of hyperparameters when training the model refers to previous studies [15]. The architecture of the model used is shown in Figure 3.

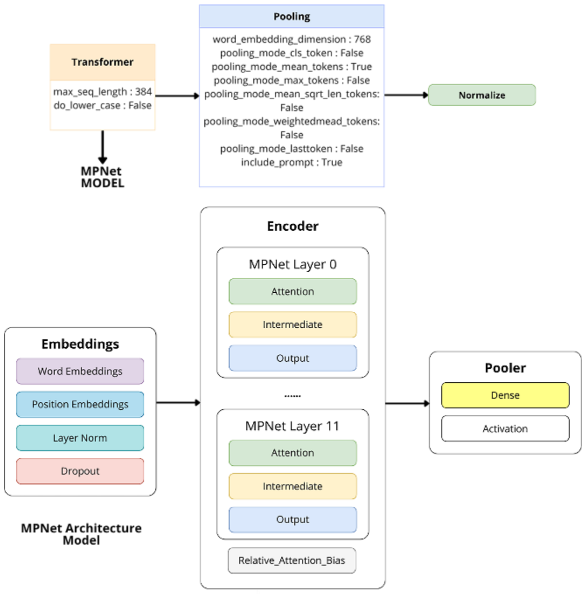

Figure 3. SBERT Model Architecture

The architecture in Figure 3 shows how the MPNet model is used as a backbone to generate text representations. The process starts from the Embeddings stage which includes word embeddings, position embeddings, layer normalization, and dropout. Then, the input goes to the MPNet Encoder Transformer which consists of 12 layers, each having attention, intermediate, and output mechanisms. After that, the resulting token representation goes through a Pooling stage, where various pooling modes can be applied such as mean or CLS-token. The final output is then normalized and can be processed by Pooler (dense + activation) for further tasks.

2) Text-to-Text Transfer Transformer (T5)

The T5 model was developed to generate CV profile summaries, with hyperparameter experiments such as learning rate, epoch, and batch size referring to previous studies [13][15]. The "t5-small" variant was chosen because it is more lightweight and efficient, in accordance with the limited computational resources. Figure 4 presents the architecture of the model used.

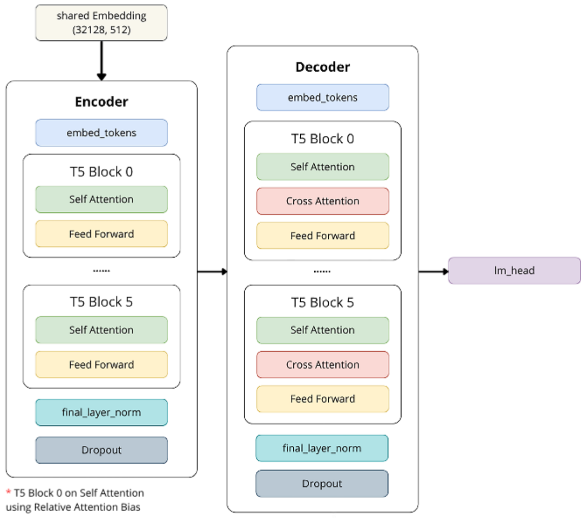

Figure 4. T5 Model Architecture

Based on Figure 4, the T5 architecture consists of a shared embedding layer shared by the encoder and decoder. The encoder consists of six T5 blocks, each having Self Attention components (with relative attention bias only in the 0th block), Feed Forward, Layer Normalization, and Dropout. The output of the encoder is sent to the decoder, which also consists of six T5 blocks, but each block has additional Cross Attention in addition to Self Attention and Feed Forward. Cross Attention allows the decoder to utilize information from the encoder. After going through all the blocks, the final result goes to lm_head to produce the final text output.

2.5. Evaluation

At this evaluation stage, there are several calculations carried out to evaluate the model that has been made at the previous stage. The calculations used to evaluate the model that has been built are shown in the following formulas:

1) F1-Score

F1-Score is a metric used to evaluate models by measuring the balance between precision and recall in the resulting predictions with respect to class imbalance. Formula (1) is used to calculate F1-Score [20] :

$$F1\text{-}Score = 2 * \frac{\textbf{Precision} * \textbf{Recall}}{\textbf{Precision} + \textbf{Recall}} \qquad (1)$$

2) ROUGE

ROUGE is a metric that is often used for the evaluation of text summarization models. It measures the similarity between the words generated by the model and its references. This calculation is performed on the evaluation of text generation models using T5. Here are some formulas used to calculate the ROUGE value:

a) ROUGE-N

ROUGE-N is a metric that measures the n-gram matching between the model result and the reference by comparing the word order generated by both. Formula (2) used to calculate ROUGE-N [21]:

$$ROUGE\text{-}N = \frac{\sum\sum \textbf{count}_{\text{match}}(\textbf{gram}_n)}{\sum\sum \textbf{count}\,(\textbf{gram}_n)} \qquad (2)$$

b) ROUGE-L

ROUGE-L is a metric that measures the Longest Common Subsequence (LCS) match between the model result and the reference, where LCS is the longest sequence of words that appear in both texts without changing their order. Formula (3) used to calculate ROUGE- L [20] :

$$ROUGE\text{-}L = \frac{(1+\beta^2)\,\textbf{Recall} * \textbf{Precision}}{\textbf{Recall} + \beta^2 * \textbf{Precision}} \qquad (3)$$

Based on the calculation of formula (3), it can be concluded that ROUGE-L combines precision and recall based on Longest Common Subsequence (LCS) to measure the similarity of word order between the model and the reference. The β value adjusts the weight between precision and recall, with $\beta > 1$ emphasizing Recall and $\beta < 1$ emphasizing Precision.

In addition to the previously described calculations, another evaluation was conducted through Confusion Matrix visualization to analyze the misclassification of the SBERT model. This matrix shows the distribution of correct and incorrect predictions in each category, making it easier to identify labels that are less accurately predicted by the model.

2.6. Deployment

This last stage, which is applying the model and system that has been built to an environment that can be accessed by users widely. The flow of the system built is shown in detail in Figure 5.

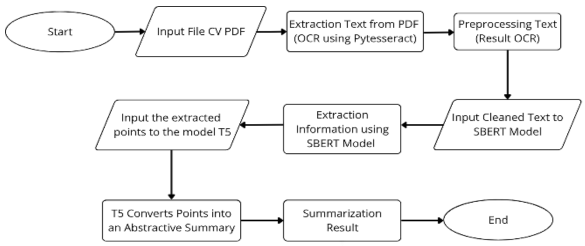

Figure 5. Flow of CV Summarization System

In Figure 5, it is shown that the flow of this system goes through several stages, starting with the input of PDF files that are processed using Optical Character Recognition (OCR) to extract text, then cleaned through the preprocessing stage. The processed text is analyzed by the SBERT model for the extraction of important information, and then abstractly summarized by the T5 model into a CV summary.

In the process of building this system, the model that has been created is first published on Hugging Face so that it can be called and used in system development. The system was developed using the Streamlit library available in Python, as it is easy to use, fast to develop, and suitable for model calls like this without requiring complex configuration. The system will then be published through Streamlit's built-in platform, streamlit.io, to make it widely accessible to users.

3. RESULT AND DISCUSSION

In this section, the results of the research process that has been carried out are presented, including the data obtained, the main findings, and information related to the objectives that have been set. The following is an explanation of the results of the research that has been carried out:

3.1 CV Information Extraction Using SBERT

Tests of information extraction from CVs were conducted using previously prepared data. This test focused on the model with the best performance compared to other models, after going through a series of experiments to obtain optimal performance results. All models in this experiment were trained for 4 epochs. Table 1 presents the hyperparameter configuration applied to each model in the information extraction experiments.

Table 1. SBERT Model Hyperparameter Configuration

Model	Split Data	Batch Size	Learning Rate
SBERT_1	80:20	8	1e-6
SBERT_2	80:20	16	2e-6
SBERT_3	70:30	8	1e-6
SBERT_4	70:30	16	2e-6

Based on Table 1, each model uses a different hyperparameter configuration. The experimental results of each model can be seen in Table 2.

Table 2. Fine-Tuning Experiments of SBERT Model

Model	Train Loss	Val Loss	Train Acc	Val Acc	F1-Score
SBERT_1	0.3204	0.3263	0.8956	0.8908	0.8866
SBERT_2	0.3305	0.3383	0.8962	0.8905	0.8864
SBERT_3	0.3483	0.3352	0.8900	0.8869	0.8822
SBERT_4	0.3425	0.3385	0.8714	0.8682	0.8770

In Table 2, the SBERT_1 model shows excellent results compared to the other three models, both in terms of accuracy and high F1-score values. Therefore, this model will be used in the system development. The test results of the model are shown in Table 3.

Table 3. Information Extraction Results of SBERT Model

Reference Information	**PERSONAL_INFORMATION** : Robert Smith ; **SKILLS** : Microsoft Excel, Great Plains, Crystal Rep, Opera, Microsoft Office, Banking, Accounts Receivable, Account Reconciliation, Accounts Payable, Human Resources, Fin, Eysite, Collections, Supervisor, Manager ; **EXPERIENCE** : Junior Accountant, Intern Vision Group Holdings ; **EXPERIENCE** : Account Reconciliation Specialist Account Opening, Closing & Maintenance Specialist Genspring Family Offices
SBERT_1 Model Results	**PERSONAL_INFORMATION**: Robert Smith Junior ; **SKILLS**: that will enable growth while utilizing my existing skills in Accounting , Commercial Collections , Customer Service , Data Management and related areas . ce , Data Management and related areas . ; **SKILLS**: . com SKILLS Microsoft Excel , Great Plains , Crystal Rep , Opera , Microsoft Office , Banking , Accounts Rec ei vable , Account Reconciliation , Accounts Pay able , Human ; **EXPERIENCE**: WORK EXPERIENCE Junior Accountant , Intern Vision Group Holdings September 2015 Present WORK EXPERIENCE Junior Accountant , Intern Vision Group Holdings September 2015 ; **EXPERIENCE**: Account Reconciliation Specialist Account Opening , Closing & Maintenance Specialist Gen sp ring Family Offices April 2015 September 2015 Responsibilities Reviewed and reconciled the assigned accounts reconciled the assigned accounts

3.2 Generative Text Using T5 and Seq2Seq

In building the generative text model, several experiments were conducted with hyperparameter adjustments. There were eight models in total, with four models each for the main approach, while the other four models were used as a comparison with the Sequence-to-Sequence (Seq2Seq) approach. Table 4 and Table 5 show the experimental process that was carried out for each approach.

Table 4. Fine-Tuning Experiments of T5 Model

Model	Split Data	Batch Size	Learning Rate	Train Loss	Val Loss
T5_1	80:20	8	2e-5	0.33450	0.24359
T5_2	80:20	4	2e-4	0.14630	0.12391
T5_3	70:30	8	2e-5	0.14900	0.10421
T5_4	70:30	4	2e-4	0.09780	0.10005

In Table 4, each model is trained for 5 epochs or iterations. It can be seen that model T5_4 has the lowest train_loss and val_loss values compared to other models. Therefore, generative text testing will be conducted using that model. However, the selection of the best model for the T5 approach will still be determined based on the evaluation results using the ROUGE metric on each model.

Table 5. Sequence-to-Sequence Model Experiment

Model	Split Data	Batch Size	Learning Rate	Train Loss	Val Loss
S2S_1	80:20	16	1e-4	2.3976	2.6334
S2S_2	80:20	32	1e-3	1.9603	2.3743
S2S_3	70:30	16	1e-4	2.4412	2.6849
S2S_4	70:30	32	1e-3	1.9395	2.4447

In Table 5, there are some differences in terms of split data, batch size, and learning rate used to train each model. Each model was trained for 35 epochs for models S2S_1 and S2S_3, and 20 epochs for models S2S_2 and S2S_4. Hyperparameter adjustments were made because the models tend to experience overfitting due to high architectural complexity, while the amount of training data available is limited. Based on the experimental results, models S2S_4 and S2S_2 were selected for further testing because they showed relatively small train_loss and val_loss values compared to other models. The testing of these models can be seen in Table 6.

Table 6. Generative Text T5 and Seq2Seq

Reference Text	Ryan Michael is a skilled professional with expertise in preparing financial reports, managing accounts departments, and conducting meetings. He holds a Master of Business Administration from San Jose State University (Jan 2019) and a B.A. in Accounting from Northeastern University (2012-2016). His experience includes roles as Chief Accountant at PETROPAL, Auditor Assistant at Aram Auditing, Accounting & Consultations, and Trainee Accountant at Hassouneh Auditing Firm. He is also a Certified Management Accountant (CMA).
Model Result T5_4	Ryan Michael is a skilled professional proficient in preparing financial reports, managing accounts departments conducting meetings, and obtaining a Master of Business Administration from San Jose State University (January 2019–January 2019), and a Finance B.A. in Accounting from Northeastern University (2012–2016). His experience includes roles as Chief Accountant at PETROPAL The Palestinian Mineral Lube Oil Co, Auditor Assistant at Aram Auditing, Accounting & Consultations, and Trainee Accountant at Hassouneh Auditing Firm. He holds a Certified Management Accountant (CMA).
Model Result S2S_2	jessica claire is a skilled professional proficient in data data data data and management, management, management, and management, and development, and a bachelor of science in computer and a bachelor of science in science from the university of art and a bachelor of science in science from the university of art….
Model Result S2S_4	jessica claire is a skilled professional proficient in data and data and data and data and data and and data and and she holds a bachelor of science in computer from the university of her experience includes roles as a as a senior analyst at the york, 2019 and a in new york, york, york, ……..

Based on Table 6, the generative text results show that model T5_4 performs much better than model S2S. The text produced by T5_4 is more coherent, informative, and close to the original reference, including important information such as name, educational background, work experience, and certification with good sentence structure. In contrast, the results from models S2S_2 and S2S_4 show a lot of word repetition, information mismatches, and sentence structure irregularities. This is likely due to the limitations of the S2S model in understanding the context of sentence length and complexity, especially when the amount of training data is limited, making it difficult for the model to produce relevant and consistent text.

3.3 Evaluation Results of Text Extraction and Generative Models

In the model evaluation stage, different metrics are calculated according to the type of task. For the span classification task used in the SBERT model, the metric used is F1-Score, which aims to measure the balance between precision and recall in predicting span labels during the training and validation process. The results of the F1-Score calculation can be seen in Table 2. In addition, a visualization of the confusion matrix is also included to show the distribution of model predictions for each label, which can be seen in Figure 6.

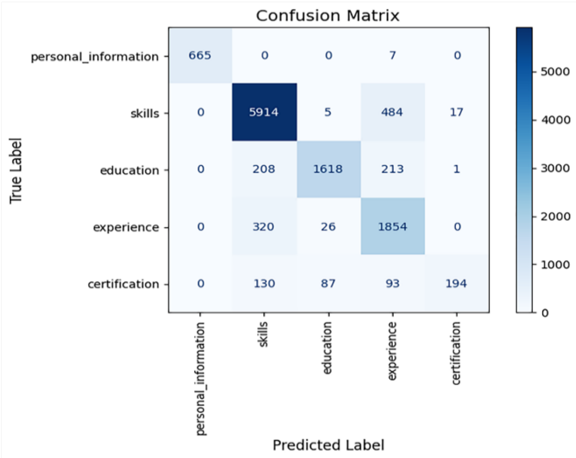

Figure 6. Confusion Matrix

Figure 6 displays the Confusion Matrix of the SBERT_1 model evaluation which shows that the model successfully classified skills (5914) and experience (1854) well. However, there are significant errors between similar categories, such as skills being frequently predicted as experience (484), and education being confused with skills (208) and experience (213). The certification category has the highest error, often predicted as skills (130), education (87), and experience (93). Personal_information had the most accurate prediction with only 7 errors. The errors occurred because some categories had similar features or overlapping contexts, making it difficult for the model to distinguish between categories.

In generative text tasks, evaluation is performed by calculating ROUGE values, specifically ROUGE-1, ROUGE-2, and ROUGE-L. Each ROUGE metric is used to measure the extent to which the generative text results match the reference. ROUGE-1 measures unigram matches, ROUGE-2 measures bigram matches, and ROUGE-L measures similarity based on Longest Common Subsequence. The metric evaluation on the generative text model is shown in Table 7.

Table 7. Evaluation of ROUGE Metrics on Generative Text Models

Model	ROUGE-1	ROUGE-2	ROUGE-L
T5_1	0.6279	0.5049	0.5835
T5_2	0.8548	0.7629	0.8304
T5_3	0.8622	0.7783	0.8413
T5_4	0.8680	0.7862	0.8483
S2S_1	0.1137	0.0431	0.1014
S2S_2	0.2153	0.0859	0.2008
S2S_3	0.1636	0.0530	0.1369
S2S_4	0.3745	0.1801	0.3272

Based on Table 7, the evaluation results using the ROUGE metric show that the model with the T5 approach has a much better performance than the S2S model. The T5_4 model scored the highest with ROUGE-1 of 0.8680, ROUGE-2 of 0.7862, and ROUGE-L of 0.8483, followed by T5_3 and T5_2 which also showed consistently high performance. In contrast, all S2S models produced low ROUGE scores, with S2S_4 being the best among the approaches but still far below the performance of the T5 model. This difference in results suggests that the T5 approach is more effective in understanding the context and points of the CV, making it more suitable for generative text tasks on the information contained in the CV.

Tests on the combined performance of the two models in a hybrid approach were conducted. The test results are shown in Table 8.

Table 8. Hybrid Model Performance Test Results

Reference Summary	Martina Lutz is a skilled professional proficient in Docker, PHP, Laravel, Python, Perl, Azure, AWS, Puppet, and SQL and NoSQL databases. She holds a B.S. in Computer Science from the University of Washington. Her experience includes roles as a Kubernetes DevOps Engineer at Ippon Technologies USA (January 2018 - Present) (Remote) and DevOps Engineer Intern at OneTrust in Kirkland, WA from June 2015 to January 2018. She holds an Amazon Web Services (AWS) certification.
Hybrid Model Results	Martina Lu tz is a skilled professional proficient in one trust deployment, deployment, and maintenance scripts, 20 installation, deployment, and maintenance scripts, 100 % mastery of deploying software through production and monitoring, supported 30 application infrastructures, and implemented appropriate environments for 30 applications. She holds a Bachelor's in Computer Science from the University of Washington. Her experience includes roles as a DevOps Engineer Intern at One Trust in Kirk land, WA (June 2015–January 2018). She holds certifications in Ama and Amazon Web Services AW S.

Based on Table 8, the comparison between the reference summary and the hybrid model results shows that although the hybrid model successfully summarizes the information, there are repetitions that affect the quality of the summary, such as errors in mentioning Martina Lutz's experience and certification. Nonetheless, the hybrid model still provides a good summary.

In addition, the hybrid model was tested on 10 documents to calculate the ROUGE score, with ROUGE-1 of 0.5497, ROUGE-2 of 0.3537, and ROUGE-L of 0.4334. These results show that the model can summarize information successfully, with a higher ROUGE-1 signifying success in matching the right words, while ROUGE-2 and ROUGE-L indicate the quality of the summary in terms of sentence order and structure.

Testing using K-Fold Cross Validation was to confirm that the selected modeling scenario is the best while minimizing bias due to split data. Because the configurations in Table 1 and Table 4 have many similarities and only difference in the split data, two scenarios from each model were selected to be tested using this method.

The test was applied to the two main models selected previously, which included SBERT and T5. Due to resource limitations, the K value used in the test was 3, as a larger value was not possible. Scenario A consists of SBERT_1 and SBERT_3 models, Scenario B of SBERT_2 and SBERT_4 models, Scenario C of T5_1 and T5_3 models, and Scenario D of T5_2 and T5_4 models. Table 9 and Table 10 are the results of K-Fold Cross Validation testing on the SBERT and T5 models.

Table 9. K-Fold Cross Validation SBERT Model

Model Scenario	Accuracy	Precision	Recall	F1-Score
Scenario A	0.8912	0.9037	0.8912	0.8866
Scenario B	0.8743	0.9109	0.8743	0.8790

Table 9 shows that in Scenario A, the model achieved an accuracy of 0.8912 and an F1-Score of 0.8866, which reflects a good balance between precision and recall. However, although Scenario B shows higher precision, the lower accuracy and recall values indicate that its overall performance is still below Scenario A.

Table 10. K-Fold Cross Validation T5 Model

Model Scenario	ROUGE-1	ROUGE-2	ROUGE-L
Scenario C	0.5857	0.3744	0.4777
Scenario D	0.7954	0.6515	0.7533

Table 10 shows that Scenario D performs better than Scenario C, with significantly higher ROUGE-1, ROUGE-2, and ROUGE-L scores. This indicates that Scenario D can generate summaries that are more accurate, informative, and relevant than Scenario C.

Therefore, the selection of SBERT_1 and T5_4 models is appropriate because the K-Fold Cross Validation results show that these configurations are effective in the performance and reliable for the modeling scenarios used.

3.4 Evaluation Results of Automatic Abstractive Summarization System

The developed system has been tested for the automatic summary process of Curriculum Vitae (CV) documents, with an average processing time of about 23 seconds per document. This time is influenced by the number of pages, because at the initial stage the Optical Character Recognition (OCR) process is carried out to extract text from the document. The more pages in the document, the longer it takes. Once the text has been extracted, the system uses the SBERT model to identify important information, which is then summarized narratively using the T5 model.

The evaluation was conducted on five CV documents in PDF format with varying number of pages. The evaluation results showed that the system was able to process the documents consistently, although there was a slight increase in duration on documents with longer pages. In addition, the quality of the summary is strongly influenced by the accuracy of text extraction through OCR. If the extraction results are less than optimal, then the summary produced also tends to be less accurate.

The developed model was published on Hugging Face to make it accessible to many users. The SBERT model, published under the name “rfahlevih/sentence-transformer-all-mpnetv2-resume-span-classifier”, is designed to classify important information in CV documents, while the T5 model, named “rfahlevih/t5-small-finetuned-resume-texgeneration”, is designed to generate narrative summaries from text that has been classified into bullet points. Both models allow users to analyze and generate CV summaries automatically.

The CV summarization system that has been built is published through the Streamlit.io platform. This platform was chosen due to its ability to display Streamlit-based applications. Figure 7 shows the CV summarization system that has been built.

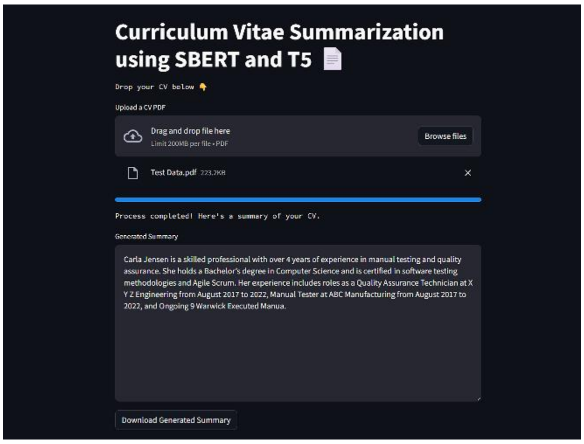

Figure 7. Curriculum Vitae Summarization System

The CV summarization system in Figure 7 works according to the flow described in Figure 5. The process starts with the user uploading the CV in PDF format. The PDF document is then extracted using OCR. Once the text has been extracted, the two models analyze and summarize the contents of the CV to produce a summary.

3.5 Research Findings

The results of this study show that the use of a combination of SBERT and T5 models can increase the effectiveness in extracting and summarizing important information from Curriculum Vitae (CV). The SBERT model shows high information classification performance with F1-score reaching 0.8866 in the best model (SBERT_1), with the ability to identify information such as skills and experience accurately, although there are still classification errors between labels that have semantic similarities. This result is in accordance with previous research [22], [23] which states that SBERT is effective for extracting information with good results.

On the generative text task, the T5 model, particularly the T5_4 variant, proved superior to the Sequence-to-Sequence (S2S) approach. This is shown by the ROUGE-1 score of 0.8680, ROUGE-2 of 0.7862 and ROUGE-L of 0.8483, which indicates that the model can produce text summaries that are close to the reference with high coherence and completeness of information. In contrast, the S2S model shows a tendency to produce repetitive and irrelevant text, which is due to its limitations in handling long contexts with limited training data.

The hybrid system combining the SBERT and T5 models successfully summarized CV documents in an average of 23 seconds per document, demonstrating efficiency that is good enough for real-world applications. However, the quality of the summarization is still highly dependent on the text extraction results from the OCR process. This suggests that the quality of the input largely determines the final performance of the system, in line with the findings of previous studies regarding the significant impact of OCR errors on NLP tasks and summary quality [24], [25]. The system evaluation showed that this approach is feasible to implement as a tool to assist job applicants in customizing their CVs for the ATS system, by providing a summarized version that better fits the format and needs of automated search.

4. CONCLUSION

This research successfully developed a Sentence-BERT (SBERT) and Text-to-Text Transfer Transformer (T5) based Curriculum Vitae (CV) summarization system that can effectively extract and summarize important information in CVs. Through a series of experiments, it was found that the SBERT_1 and T5_4 models gave the best results on each task, with high evaluation values on F1-Score and ROUGE metrics. The integration of these two models in the hybrid system proved to be able to summarize CVs in a format that is more relevant and easily processed by automated selection systems such as the Applicant Tracking System (ATS). The evaluation results showed that the hybrid system obtained ROUGE-1 scores of 0.5497, ROUGE-2 scores of 0.3537, and ROUGE-L scores of 0.4334, indicating a good summarization performance in retaining key information.

The developed system is not only capable of producing accurate and informative summaries, but also efficient in terms of processing time. The system has the potential to assist job applicants in customizing their CVs to better fit the automated hiring criteria commonly used by companies.

Future research can improve the quality of Optical Character Recognition (OCR) to obtain more accurate text extraction results, especially on CVs with complex formats. In addition, the addition of training data from various industry backgrounds can help improve the generalization of the model. The system can also be equipped with an automatic adjustment feature based on the job description, so that the summary produced is more relevant and in accordance with the needs of the ATS system.

Acknowledgment

All praise is due to Allah SWT for His grace and guidance, which enabled us to complete this research journal. The authors would appreciate the supervisors for their guidance, direction, and support during the research process. In addition, the author would like to thank the entire academic community of the Informatics Engineering Department of UIN Sunan Gunung Djati Bandung for all the assistance provided, so that this research can be completed properly.

5. REFERENCE

[1] D. F. Mujtaba and N. R. Mahapatra, "Fairness in AI-Driven Recruitment: Challenges, Metrics, Methods, and Future Directions," May 2024.

[2] P. Horodyski, "Recruiter's perception of artificial intelligence (AI)-based tools in recruitment," *Computers in Human Behavior Reports*, vol. 10, p. 100298, May 2023, doi: 10.1016/j.chbr.2023.100298.

[3] G. Navarro, "Fair and Ethical Resume Screening: Enhancing ATS with JustScreen the ResumeScreeningApp," Journal of Information Technology, Cybersecurity, and Artificial Intelligence, vol. 2, no. 1, pp. 1–7, Jan. 2025, doi: 10.70715/jitcai.2024.v2.i1.001.

[4] P. R. Chavan, Y. Chandurkar, A. Tidake, G. Lavankar, S. Gaikwad, and R. Chavan, "Enhancing recruitment efficiency: An advanced Applicant Tracking System (ATS)," Industrial Management Advances, vol. 2, no. 1, p. 6373, Jul. 2024, doi: 10.59429/ima.v2i1.6373.

[5] A. Mukhtar and M. Masradin, "Bagaimana Teknologi Era 4.0 Menerapkan Rekrutmen? (Kajian Manajemen Sumber Daya Manusia)," Paraduta : Jurnal Ekonomi dan Ilmu-Ilmu Sosial, vol. 1, no. 2, pp. 77–89, Sep. 2023, doi: 10.56630/paraduta.v1i2.479.

[6] R. V. K. Bevara et al., "Resume2Vec: Transforming Applicant Tracking Systems with Intelligent Resume Embeddings for Precise Candidate Matching," Electronics (Basel), vol. 14, no. 4, p. 794, Feb. 2025, doi: 10.3390/electronics14040794.

[7] S. Bharadwaj, R. Varun, P. S. Aditya, M. Nikhil, and G. C. Babu, "Resume Screening using NLP and LSTM," in 2022 International Conference on Inventive Computation Technologies (ICICT), IEEE, Jul. 2022, pp. 238–241. doi: 10.1109/ICICT54344.2022.9850889.

[8] Dimas Bagus Susanto and Said Hamzali, "The Role of Technology in Improving the Effectiveness of Employee Recruitment and Selection," Journal of Economic Education and Entrepreneurship Studies, vol. 5, no. 3, pp. 421–434, Aug. 2024, doi: 10.62794/je3s.v5i3.3800.

[9] M. Lavanchy, P. Reichert, J. Narayanan, and K. Savani, "Applicants' Fairness Perceptions of Algorithm-Driven Hiring Procedures," Journal of Business Ethics, vol. 188, no. 1, pp. 125–150, Nov. 2023, doi: 10.1007/s10551-022-05320-w.

[10] J. Opitz and A. Frank, "SBERT studies Meaning Representations: Decomposing Sentence Embeddings into Explainable Semantic Features," in Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Stroudsburg, PA, USA: Association for Computational Linguistics, 2022, pp. 625–638. doi: 10.18653/v1/2022.aacl-main.48.

[11] M. Yani, N. S. Khodijah, R. Rendi, and M. Mustamiin, "Aplikasi Peringkas Teks Bahasa Indonesia Menggunakan Model Text-to-Text Transfer Transformer (T5)," IKRA-ITH Informatika : Jurnal Komputer dan Informatika, vol. 9, no. 2, pp. 78–86, Oct. 2024, doi: 10.37817/ikraith-informatika.v9i2.4392.

[12] A. Deshmukh and A. Raut, "Enhanced Resume Screening for Smart Hiring Using Sentence-Bidirectional Encoder Representations from Transformers (S-BERT)," International Journal of Advanced Computer Science and Applications, vol. 15, no. 8, 2024, doi: 10.14569/IJACSA.2024.0150825.

[13] G. E. Abdul, I. A. Ali, and C. Megha, "Fine-Tuned T5 for Abstractive Summarization," International Journal of Performability Engineering, vol. 17, no. 10, p. 900, 2021, doi: 10.23940/ijpe.21.10.p8.900906.

[14] D. S., S. N., J. Andrew, and M. Mazzara, "Unified extractive-abstractive summarization: a hybrid approach utilizing BERT and transformer models for enhanced document summarization," PeerJ Comput Sci, vol. 10, p. e2424, Nov. 2024, doi: 10.7717/peerj-cs.2424.

[15] P. Vajdecka, V. Svatek, and M. Vita, "A Novel Approach to Abstractive Summarization Based on LOF, Sentence-BERT and T5 – with Fact Checking Use Case," 2023. doi: 10.2139/ssrn.4493592.

[16] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," Procedia Comput Sci, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[17] U. Kannengiesser and J. S. Gero, "Modelling the Design of Models: An Example Using CRISP-

DM," Proceedings of the Design Society, vol. 3, pp. 2705–2714, Jul. 2023, doi: 10.1017/pds.2023.271.

[18] M. Mlaweh, "Datasets Annotated NER PDF Resumes," Hugging Face. Accessed: Mar. 01, 2025. [Online]. Available: https://huggingface.co/datasets/Mehyaar/Annotated_NER_PDF_Resumes.

[19] Y. Khalil, "Resumes-Images-Datasets," Kaggle. Accessed: Mar. 01, 2025. [Online]. Available: https://www.kaggle.com/datasets/youssefkhalil/resumes-images-datasets.

[20] A. Ibrahim, M. Alfonse, and M. Aref, "A Systematic Review on Text Summarization of Medical Research Articles," International Journal of Intelligent Computing and Information Sciences, vol. 23, no. 2, pp. 50–61, Jun. 2023, doi: 10.21608/ijicis.2023.190004.1252.

[21] Z. Yao, A. Chen, and H. Xie, "Chinese long text summarization using improved sequence-to-sequence lstm," J Phys Conf Ser, vol. 1550, no. 3, p. 032162, May 2020, doi: 10.1088/1742-6596/1550/3/032162.

[22] Y. Liu and S. Li, "AutoIE: An Automated Framework for Information Extraction from Scientific Literature," Jan. 2024.

[23] M. Gebauer, F. Maschhur, N. Leschke, E. Grünewald, and F. Pallas, "A 'Human-in-the-Loop' approach for Information Extraction from Privacy Policies under Data Scarcity," in 2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), IEEE, Jul. 2023, pp. 76–83. doi: 10.1109/EuroSPW59978.2023.00014.

[24] D. van Strien, K. Beelen, M. Ardanuy, K. Hosseini, B. McGillivray, and G. Colavizza, "Assessing the Impact of OCR Quality on Downstream NLP Tasks," in Proceedings of the 12th International Conference on Agents and Artificial Intelligence, SCITEPRESS - Science and Technology Publications, 2020, pp. 484–496. doi: 10.5220/0009169004840496.

[25] P. P. M. Chau, S. Bakkali, and A. Doucet, "DocSum: Domain-Adaptive Pre-training for Document Abstractive Summarization," Dec. 2024.