

Laptop Recommendation Intelligent Virtual Assistant using Recurrent Neural Network with RPA for Data Scraping

Steven Tjayadi
Faculty of Information Technology
Tarumanagara University
West of Jakarta, Indonesia
steven.535180085@stu.untar.ac.id

Viny Christanti Mawardi
Faculty of Information Technology
Tarumanagara University
West of Jakarta, Indonesia
viny@untar.ac.id

Abstract— Virtual Assistant is a job that helps receive support services for customers online or remotely. Virtual Assistant can be named as a chatbot in an Artificial Intelligence background. The Virtual Assistant we used is for Laptop Recommendation. Laptops are modern, multi-functional computing devices that can be used just like computers. For Laptop Recommendation, we use Generative text Virtual Assistant with Recurrent Neural Network. It uses an access link to connect to a laptop on the website itself so the user can see the full description of the laptop. Before making a text, it used a UiPath for gathering the data for training. In the Recurrent Neural Network method, encoder-decoder is used to predict answers based on question input. The accuracy for these methods is 96% and can be accessed at <http://valaptop.labirariset.com>. We also tested respondents to find out the results of direct responses from users. Users who use non-standard sentences produce false responses.

Keywords— Data, laptop, Recurrent Neural Network, UiPath, Virtual Assistant

I. INTRODUCTION

Laptop can be used as the same alternative as a computer. Usually, people use a laptop for installing an application like games, offices, design, and other specific applications. Since in the application the user in the laptop downloads sometimes always update based on user experience and specification of a laptop, that makes a laptop production are even more expensive and more advanced like laptop gaming also the other brand is even more variance like MSI stealth, Acer predator, and the others. Sometimes not all people can understand the specification and standards of a laptop because of the technology increasingly so people need a recommendation system to choose a laptop for the long term [1]. There are 2 types of Recommendation Systems, the first one is using the Machine Learning method and the second one is Virtual Assistant with Chatbot. With Machine Learning methods there were using K-Nearest Neighbors by Raharja [2] and using TOPSIS by Fei Lie [3]. Virtual Assistant is a good and a new thing to make recommendations smoothly and also help people who do not much about the laptop [4].

Virtual Assistant is a job that helps receive support services online or remotely (Alternatives as Chatbot). Virtual Assistants really help the development of technology, especially in the current era of the internet so that it can be done anywhere and anytime and there is no need to bring new employees to do work like customer service. Virtual Assistants are very helpful for businesses, especially in recommendation services systems. Virtual Assistants use the technology of Artificial Intelligence (AI) and Natural

Language Processing based on of the development of AI and technology [5].

Virtual Assistant (Chatbot) has 2 types, there were Generative Chatbot and Retrieval Chatbot. Generative Chatbots use a machine learning model to predict the answer based on the question. Retrieval Chatbot uses a query similarity to predict the answer based on the question [6]. In this topic, it uses Generative Chatbot [7].

When using Virtual Assistant, it needs a lot of data to make a Virtual Assistant run optimally. When collecting the data, most developers do not really want to repeat it like just copying and pasting to collect the data one by one. Also, they want to search for the data that has already been built from the internet and did not want to make new data. So, to solve this needs Robotic Process Automation (RPA) to make data collection easier [8]. RPA really helps users to get the data collected effectively by using the technology UiPath. UiPath was a user-friendly technology application in RPA and a free trial with no programming language background. RPA really helps not only in data collecting, but it also has many functions of automation like email automation, excel automation, and many else. Also, RPA really helps people for doing other tasks instead of doing repetitive tasks like input and output data.

II. METHODS

A. Robotic Process Automation (RPA)

Robotic Process Automation (RPA) is a method used to automate repetitive tasks like collecting data. RPA has so many successful technologies in many industries and can help with many tasks. The example of tasks that can be done in RPA are [9]:

- Scraping data from the website
- CRUD (Create, Read, Update, Delete) Data files and Folder
- Connecting to Website, Database, and application
- Extracting data from documents such as images and pdf files
- Calculation in system data

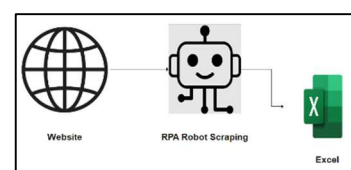


Fig. 1. Web Scraping With Robotic Process Automation

We can see example of illustration of web scraping with RPA at Figure 1. The Technology used for RPA is UiPath Studio. UiPath Studio has many tools to make automate easily. In UiPath Studio, there are many functions and libraries that can be used for automating tasks.

The UiPath Studio is used for this topic to collect the laptop and dialogue data with UiPath. When using UiPath, it used the Data Scrapping activity. Then, select the data and website you want to collect so the UiPath will automatically read the data based on the same pattern. At Figure 2 we can see UiPath need the keyword so the system can automatically scrap the data from the web.

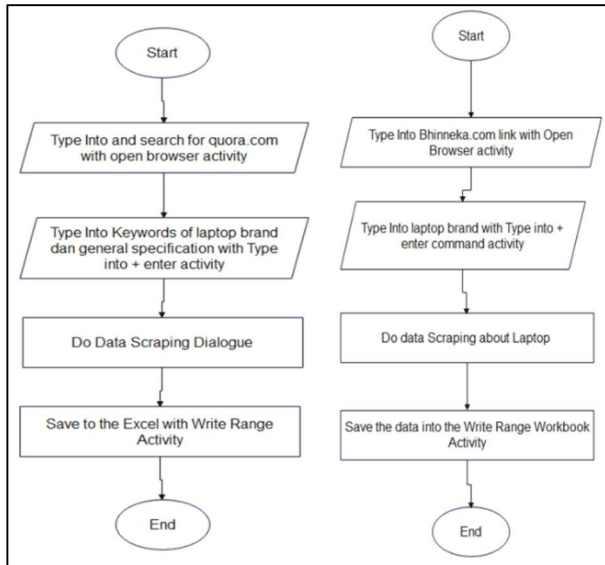


Fig. 2. Scraping with RPA (UiPath)

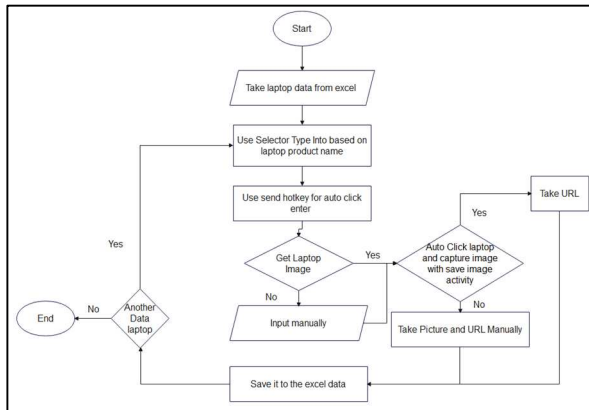


Fig. 3. Preprocessing Laptop Data with UiPath

At Figure 3 we can see preprocessing to get image from the web based on the list data that already collected from previous step. Finally, at Figure 4 we translate the data if we get text in English. All the answer of chatbot will be display in Bahasa.

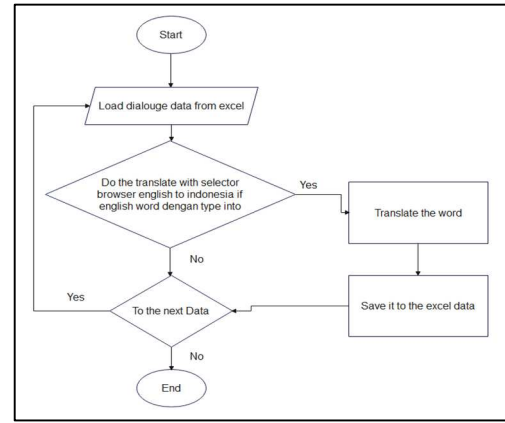


Fig. 4. Preprocessing Dialogue Data with UiPath

After collecting the data, it will be saved to an excel (Dialogue data) or csv (laptop data) file to get the train with Recurrent Neural Network. For the laptop data, it will send to the database file with SQLite framework to connect to the Backend of the website administration.

B. Recurrent Neural Network (RNN)

The method used to train the model of the Virtual Assistant was using a Recurrent Neural Network (RNN). This method also uses collaboration with the Long Short-Term Memory (LSTM) cell which is one of the optimal methods to do the Deep Learning model, especially in the Virtual Assistant (Chatbot) model. This method uses an encoder-decoder model which makes a Virtual Assistant make a prediction answer text based on a question-and-answer trained dataset. Encoder-Decoder pattern can be seen in Figure 5.

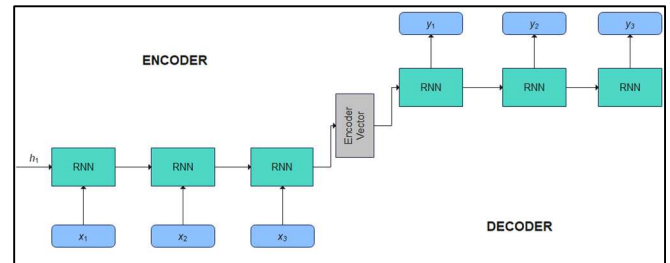


Fig. 5. Encoder-Decoder RNN

To do the encoder-decoder methods with RNN, there are several ways to understand its type [10].

- Encoder: It will read as an input vector in a question. It will train and read as a previous hidden state (h_t) from the input vector (x_t).

$$h_t = f(h_{t-1} + x_t) \quad (1)$$

- Decoder: It will read from the question and the last answer from the encoder answer.

$$h_t = f(h_{t-1}, y_{t-1}, c) \quad (2)$$

After doing the Encoder-Decoder with RNN it will get the prediction result with the (y_t) symbol and will follow the softmax method and the Weight of the SoftMax (S).

$$y_t = softmax(W^S h_t) \quad (3)$$

C. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) can take cover the Recurrent Neural Network vanishing gradient. LSTM used a cell that connected inside two encoders and decoders of RNN so the process of counting RNN results will get a smooth result. Inside the cell of LSTM, they have multiple cells and gates used to store and forget the unnecessary information of text so that can focus on optimal training. LSTM Cell can be seen in Figure 6 [11].

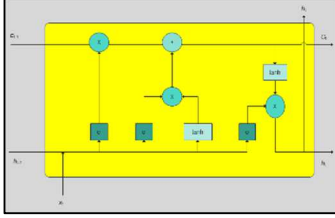


Fig. 6. LSTM Cell

LSTM Cell can be implemented inside the RNN model in Figure 6 by this default following variable and description method: [11]

$$\sigma = \frac{1}{1+e^{-x}} \quad (4)$$

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (5)$$

$$I_t = \sigma(W_i * [h_{t-1}, x_t]) \quad (6)$$

$$O_t = \sigma(W_o * [h_{t-1}, x_t]) \quad (7)$$

$$C_{t-1} = \tanh(W_c * [h_{t-1}, x_t]) \quad (8)$$

$$C_t = h_{t-1} * f_t + I_t * C_{t-1} \quad (9)$$

$$h_t = O_t * \tanh(C_t) \quad (10)$$

Where σ is Sigmoid, I_t is Input Cell, O_t is Output Cell, C_{t-1} is New Candidate Cell, C_t is Current Candidate Cell, h_t is Hidden State Cell and x_t is Input from vector (Question).

D. Virtual Assistant (Chatbot) Type

The Virtual Assistant used in Recurrent Neural Network (RNN) was a Generative text. Generative Virtual Assistant uses natural language understanding and a deep learning model to make a prediction with text based on the question [12]. It used tokenizer and padding to make Virtual Assistant can create and train the question based on the long and short sentence pattern.

E. SQLite

SQLite was used to match between the data and the trained data in the Recurrent Neural Network. It uses laptop data collected with RPA and matches the laptop name with SQL query so the trained data will automatically make a URL link for the answer of the laptop name to connect to the selected laptop based on the trained data. SQLite can upload the laptop data with .csv format easily and can directly call in the Backend website like Flask. In Flask, it uses a pip install package to make a function of SQLite and match the trained data laptop connect into laptop website [13].

F. Python Flask

Flask uses Python with pip installation for the administration of the Backend Website. It also uses a specific file default folder in HTML, CSS, Javascript, and Images with a specific folder and link to connect with the Flask. In HTML files it uses the "Templates" folder name and the remaining use static and uses each subfolder CSS, img, and JS. Then they will connect with the syntax "url_for" in the Html file based on the location of the subfolder to connect. Finally, in Python files will use the route and return render_template based on the name of Html files to connect each other with the front-end and back-end [13].

III. RESULT AND DISCUSSION

Based on the Flowchart at figure 7, there are 2 ways and data need to connect to the website. That is from laptop data and dialogue data.

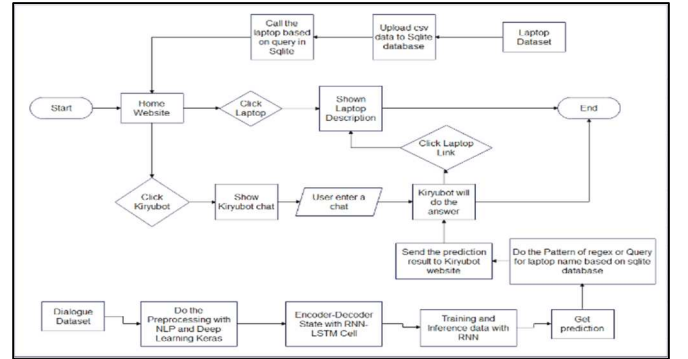


Fig. 7. Flowchart Design Result

Based on separated ways, the detail of the step of the laptop data and dialogue data were:

- 1) Laptop Data
 - a) Upload csv laptop data to the database SQLite
 - b) Connect to SQLite and save the laptop data file with the .db format
 - c) Connect SQLite in Flask with pip package pysqlite3 installation
 - d) Make a call with SQL query in Flask that connected from SQLite based on HTML variable format and matches with the event button data in HTML to connect the laptop data
- 2) Dialogue Data
 - a) Preprocessing the dialogue data with Natural Language Processing (Regex) and Deep Learning (Keras Tokenizer) method
 - b) Use an Encoder-Decoder RNN with the collaboration of LSTM cell
 - c) Train and inference with Recurrent Neural Network
 - d) Get the prediction of the model
 - e) Do the pattern with Query in Python Flask to make a connection with the model and the laptop data if matches

- f) Send the prediction result to the Virtual Assistant Website

Based on a collection laptop data and dialogue data it will use RPA UiPath training data and will get a result in RPA testing. Laptop data will get a result in Website Testing and Dialogue data will get a result in Recurrent Neural Network testing. After a website is done, it will share to the viewers to get a User Accepted Test (UAT) and feedback. The result of the UAT testing will be shown in Virtual Assistant Testing.

A. RPA Testing

Based on RPA Testing with UiPath, there are 2 types that can also get good automation with RPA:

- Laptop data, it can scrap the laptop name, price, and URL link for the laptop and automatically capture and save the laptop product picture based on its name. It used the subfolder "GambarLaptop" with "\\" and automatically put the laptop picture with the file name based on the laptop name. Examples of laptop data scraping are in Table 1.

TABEL I. LAPTOP DATA

Laptop Name	ACER PREDATOR TRITON 500 PT515-51-713V
Price	Rp 35.999.000
Laptop Picture	GambarLaptop\ACER Predator Triton 500 PT515-51-713V.png
Brand	Acer
Processor	Intel Core i7-9750H
RAM	16 GB
SSD	512 GB SSD
Graphics	NVIDIA GEFORCE RTX 2060
Size	15.6"
Resolution	FHD
Description	Acer Predator Triton 500 punya desain inovatif dan spesifikasi tinggi. Dimensi laptop yang ringkas ini membuat Predator Triton 500 mudah dimasukkan ke dalam tas ransel. Predator Triton 500 punya berat kurang lebih 2.1 kg dan tebal hanya 17.9 mm. Spesifikasi yang dimiliki laptop ini sangat gahar dan mumpuni untuk memainkan game dalam setingan grafis rata kanan. Tampilan laptop gaming ini semakin gahar dengan balutan warna Black Metal. Layar yang besar menambah keseruan saat main game favorit.
URL	https://www.bhinneka.com/acer-predator-triton-500-pt515-51-713v-nhq50sn002-sku3324066520

- Dialogue data, it can scrap dialogue question and answer in Quora and can translate language English to the Indonesian language in Question-and-answer data.

In Table 2, there is an example of the words which need to be automated translated from English to Indonesia with UiPath.

TABEL II. TRANSLATED QUESTION ENGLISH TO INDONESIA

No	Question	Translated Question
1	Is MSI is a good brand?	<i>Apakah MSI merek laptop yang bagus?</i>
2	Is MSI better than Acer?	<i>Apakah MSI lebih baik dari Acer?</i>
3	Is an Intel i7 always faster than an i5?	<i>Apakah Intel i7 selalu lebih cepat daripada i5?</i>
4	What is the best laptop for freelancer?	<i>Laptop apa yang terbaik untuk freelancer?</i>
5	What are the best laptops for Computer Science and Engineering students?	<i>Apa laptop terbaik yang bisa saya beli sebagai mahasiswa Computer Science dan Teknik?</i>

Based on RPA testing with UiPath, there are 2 types of weaknesses, first cannot scrap a laptop specification perfectly and sometimes users need to click the laptop product if the automation automatically stops after searching.

Also, the data collection with RPA for each species in Laptop data and Dialogue data. There are 63 data for laptops which consists from 5 brand Laptop namely, MSI, Acer, Asus, Hp, and Lenovo. We can see the number of data each brand at table III.

TABEL III. LAPTOP DATA COLLECTION

No	Brand	Data
1	ASUS	12
2	MSI	13
3	Acer	7
4	HP	21
5	Lenovo	10

We can see at table IV, dialogue data collection that consist of 400 dialogues and categorized in 3 categories.

TABEL IV. DIALOGUE DATA COLLECTION

No	Topic	Total
1	Laptop Product recommendation	200
2	Laptop Specification Recommendation	100
3	System Requirement (Specification)	100

B. Recurrent Neural Network (RNN) Model Testing

In Recurrent Neural Network (RNN) Testing, it will use the Recurrent Neural Network based on the encoder-decoder method with LSTM cell. There are 2 types of models of the virtual assistant status of the recommendation laptop which are accurate and not accurate. The model testing based on the status of recommendation laptop training can be seen in Table V.

TABEL V. RNN EXAMPLE MODEL TESTING

No.	Model Testing		
	Question	Answer	Status
1	Apakah laptop Acer Nitro 5 layak dibeli?	Iya Silahkan	Accurate
2	Lebih baik Lenovo atau Acer?	Disarankan Lenovo dengan Acer disarankan Lenovo untuk kualitas lebih baik	Accurate
3	Manakah laptop gaming MSI terbaik?	MSI GF63 Thin akan membantumu	Accurate
4	Apakah MSI merek laptop yang bagus?	Ya msi adalah merek perusahaan yang bagus dan mesin laptop yang bagus dengan harga yang sangat wajar	Accurate
5	Fungsi VGA?	Wawasan tentang ram ada	Not Accurate

After collected 400 data dialogues, we train the system until epoch score 75. The accuracy result can be seen at Table VI. In this epoch the model has been stable and has a loss value of 0.0614. With this value, the RNN model with 400 training data can be used for virtual assistants.

TABEL VI. TESTING RESULT RNN MODEL

Percentage (Train-Test)	Accuracy	Precision	Recall	F1-Score
80 - 20	96,56	98,99	92,79	95,79

C. Website Testing

The result of model combined by using the new website to place a laptop and a chat. The nickname of the website is **Micro-Virtual Laptop (MVL)** and the virtual assistant is named **Kiryubot**. Adopting Kusnadi's research, this study also uses Bhinneka as a site that is linked to displaying laptop products [14]. A laptop product from Kiryubot could also be connected to another website like Bhinneka and put on the laptop description screen (Figure 10) so the user can automatically place a new tab to see the price and details about a laptop. There is also a connection for a Virtual Assistant (**Kiryubot**) in the right-bottom (floated) of the Home Screen (Figure 8) which can help answer based on the RNN model.



Fig. 8. Home Screen

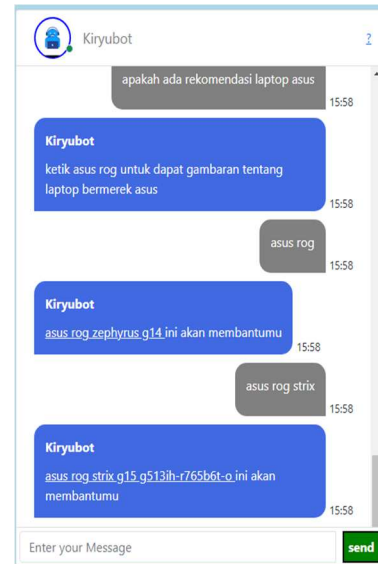


Fig. 9. Virtual Assistant Chat Screen

Also, there is a chance of the predicted of laptop answer (Figure 9) which give the target site by link if based on the laptop product name so the user can see the description and specification about laptop recommendation with a new tab. It used SQL for searching a laptop on the website based on database and query syntax. The website model result can be seen in Figure 8 until Figure 10.

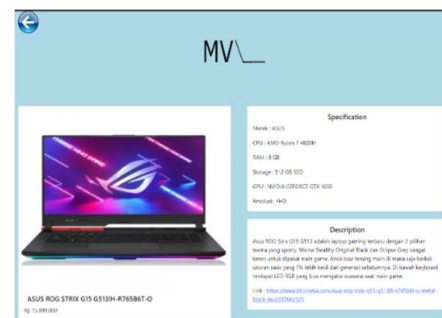


Fig. 10. Laptop Description Screen

D. Virtual Assistant Testing

The model testing of the Virtual Assistant was 75 % make a good sentence and fit the data. The predicted of the Virtual Assistant and laptop recommendation link are linked correctly.

However, there is also a weakness in the virtual assistant model chat. There are:

- User must use the full text based on Indonesian and not slang language or abbreviated words.
- The exception of the laptop user can type the code of the laptop like "Asus Notebook", "Acer Predator", and so on to get the link of the laptop in Virtual Assistant chat.
- Users cannot use any punctuation words in question chat except "?" and "-" like for example "kira-kira" and the code of the laptop name (Asus rog strix g15 g513ih-r765b6t-o)

This testing has been tested by 30 respondents. It used a survey with google form and test the virtual assistant website virtually with the website URL given. The result of the survey can be seen in with pie chart in Figure 11.

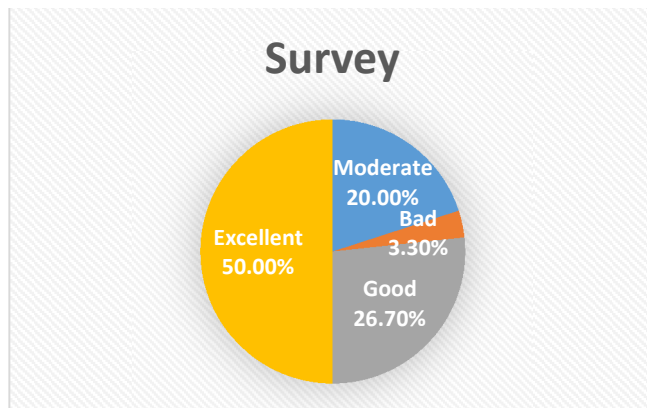


Fig. 11. Virtual Assistant Survey of 30 Respondent

Based on the result of the survey of the Virtual Assistant model, 23 out of 30 Respondents really trust and like this virtual assistant with a total of 50 % Excellent and 26,7 % good.

IV. CONCLUSION

Based on the result of the Website testing and Virtual Assistant Testing with the automation data scraping, the training of the recommendation laptop:

- Gathering data with UiPath is efficient and accurate for many specific repetitive data and file names.
- The accuracy of the Recurrent Neural Network model was 96 % with a precision of 99 %, recall of 93 %, F1-Score 95 %, and Categorical Cross-Entropy loss of 0,22.
- Virtual Assistant chat need to use formal or good Indonesian word to make a good prediction.
- User can type a brand on a laptop code like “Asus”, “Asus Rog” and the other else in the virtual assistant chat.
- Recurrent Neural Network method really gets a good result and can help predict the great result with the help of Chatbot type.
- In the future the laptop virtual assistant, will need a complex Deep Learning of the Chatbot type method like Bidirectional LSTM or use Retrieval type can help cover the weakness of the Recurrent Neural Network Virtual Assistant Chatbot and get a multiple variety

prediction of the answer with the help of Robotic Process Automation.

REFERENCES

- [1] M. S. Ayundhita, Z. K. A. Baizal, and Y. Sibaroni, "Ontology-based conversational recommender system for recommending laptop." In *Journal of Physics: Conference Series*, vol. 1192, no. 1, p. 012020. IOP Publishing, 2019.
- [2] C. A. Rahardja, T. Juardi, and H. Agung, "Implementasi algoritma K-Nearest Neighbor pada website rekomendasi laptop." *Jurnal Buana Informatika* 10, no. 1 (2019): pp. 75–84.
- [3] F. Lie and T. T. Suryosuseno, "Sistem pendukung keputusan Pemilihan laptop menggunakan metode Topsis." *CAHAYATECH* 7(2), 2018, pp. 119–134.
- [4] B. S. Darshan, S. Ajay, C. Akshatha, V. Aishwarya, and S. G. Shilpa, "Virtual assistant-based recommendation system." *Int. J. Adv. Res. IdeasInnov. Technol.* 5(3), 2019, pp. 1191–1194.
- [5] E. Yuniar and H. Purnomo, "Implementasi chatbot Alitta" asisten virtual dari balittas sebagai pusat informasi di balittas." *Antivirus: Jurnal Ilmiah Teknik Informatika* 13(1), 2019, pp. 24–35.
- [6] A. Ciayandi, V. C. Mawardi, and J. Hendryli, "Retrieval based chatbot on tarumanagara university with Multilayer Perceptron." In *IOP Conference Series: Materials Science and Engineering*, vol. 1007, no. 1, p. 012146. IOP Publishing, 2020.
- [7] S. A. Sheikh, V. Tiwari, and S. Singhal, "Generative model chatbot for human resource using deep learning." In *2019 International Conference on Data Science and Engineering (ICDSE)*, pp. 126–132. IEEE, 2019.
- [8] J. Ribeiro, R. Lima, T. Eckhardt, and S. Paiva, "Robotic process automation and artificial intelligence in industry 4.0—a literature review." *Procedia Computer Science* 181, 2021, pp. 51–58.
- [9] M. Florentina, "Web data extraction with robot process automation. study on LinkedIn web scraping using Uipath Studio." *Annals of Constantin Brancusi University of Targu-Jiu. Engineering Series* 1, 2020.
- [10] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation." *arXiv preprint arXiv:1406.1078*, 2014.
- [11] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging." *arXiv preprint arXiv:1508.01991*, 2015.
- [12] Z. Yin, K. H. Chang, and R. Zhang, "Deepprobe: Information directed sequence understanding and chatbot design via recurrent neural networks." In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2131–2139, 2017.
- [13] S. Suraya and M. Sholeh, "Designing and implementing a database for thesis data management by using the python Flask Framework." *International Journal of Engineering, Science and Information Technology* 2(1), 2022, pp. 9–14.
- [14] A. Kusnadi and E. Kurniawan, "Implementation of topsis method in web-based system recommendations for student's laptop selection (case study: Bhinneka. com)." *IJNMT (International Journal of New Media Technology)* 4(1), 2017, pp. 42–45.