Virtual Route Guide Chatbot Based on Random Forest Classifier

Puspa Miladin Nuraida Safitri A. Basid[1], Fajar Rohman Hariri[2], Fresy Nugroho[3], Ajib Hanani[4], Firman Jati Pamungkas[5]
Informatics Engineering, Faculty of Science and Technology,
Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia[1, 2, 3, 4]
Library and Information Science, Faculty of Science and Technology,
Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia[5]

***Abstract*—Improvements in the quality of tourism services and the number of human resources will affect the quality of social services and information services provided to foreign tourists, thereby enhancing the quality of services offered regarding tourist destination information in the Malang Raya area. Considering the urgency of foreign tourists in obtaining information related to directions, routes, and access roads to their desired tourist destinations, especially in East Java, due to limited data from the government agencies handling the tourism sector, as well as the difficulty in communication with residents who may not understand what is being communicated by foreign tourists. Therefore, the need for an interactive chatbot to assist in obtaining routes and access information to the desired tourist destinations will facilitate foreign tourists. To improve the accuracy of the chatbot's ability to answer sentence selection, the use of artificial intelligence, specifically the Random Forest Classifier, is necessary. This study obtained the highest accuracy value using a tree quantity of 200, a maximum tree depth of 20, and a minimum sample split of 5. Using these quantities resulted in an accuracy of 95.88%, precision of 96.29%, recall of 96.03%, and f-measure of 96.16%.**

Keywords—Tourism; chatbot; artificial intelligence; random forest classifier

I. Introduction

Tourism is a dynamic activity involving many people and stimulating various business sectors. In the current era of globalization, the tourism sector will become the main driver of the world economy and global industry. Tourism will provide significant revenue for regions aware of its potential in the tourism sector [1]. Thus, tourism has become an integral part of human life. The tourism sector has led several regions in Indonesia to develop tourism as their distinctive feature.

One area with great tourism potential is Malang, located in the East Java province of Indonesia. Malang Raya has three administrative areas: Malang Regency, Malang City, and Batu City. Malang Raya is one of the leading tourist destinations in East Java and Indonesia. According to statistical data compiled by the Ministry of Tourism and Creative Economy in 2022, the total number of international tourists visiting Malang Raya reached six million [2].

The Malang Raya region offers many beautiful tourist destinations unique to other areas in Indonesia. Malang Raya offers various tourism categories, from natural to manufactured attractions. According to data from the Malang Regency's One Data program, there are 16 village tourism objects, 106 natural tourist attractions, 49 cultural tourist attractions, and 24 manufactured tourist attractions in the Malang Regency area [3]. According to data from the Central Statistics Agency (BPS) of Batu City, there are ten manufactured tourist attractions, 12 village tourism objects, five natural tourist attractions, five souvenir tourist attractions, and one religious tourist attraction in the Batu City area [4]. According to data from Malang's One Data program, there are 16 cultural tourist attractions, two historical tourist attractions, four religious tourist attractions, one educational tourist attraction, 2015 culinary tourist attractions, 12 shopping tourist attractions, and 20 manufactured tourist attractions in the Malang City area [5].

The abundance of tourist destinations in the Malang Raya region provides many options for international tourists. However, on the other hand, with the increasing number of tourist destinations, issues arise regarding information about the destinations to be visited. International tourists require comprehensive information about routes, directions, and ways to guide their travel, but not all available information from print media, television, the Internet, and other sources can meet these needs [6]. Another issue in the tourism sector is the low quality of service and quantity of human resources in the tourism industry worldwide. This aspect needs special attention in efforts to improve the tourism sector in the Malang Raya region. The quality of tourism services and the quantity of human resources are among the standards that will be compared to achieve tourist satisfaction [7]. Improvements in the quality of tourism services and the number of human resources will affect the quality of social service and information services provided to international tourists, thereby enhancing the quality of services related to tourist destination information in the Malang Raya region.

Another issue encountered is the current condition of the official websites of the Department of Culture and Tourism of Malang Regency, Malang City, and Batu City, which only provide brief information about tourist attractions and lacks interactive question-and-answer features that can guide foreign tourists in terms of directions, routes, and roads to the destinations. As a result, foreign tourists have to search for routes or access on their own through several stages. The lack of digital information services like this can lead to inefficiency and ineffectiveness in obtaining access information or routes to selected tourist destinations by foreign tourists [7].

Another issue is the poor communication between residents and foreign tourists. This is due to the lack of knowledge, which prevents residents from understanding what foreign tourists communicate [8]. This has become a significant concern for the government, notably the Tourism Sector Institution in the Malang Raya area. Considering the urgency and the difficulty faced by foreign tourists in obtaining information regarding directions, routes, and access to tourist destinations, especially in East Java, due to limited information from the government agencies handling the tourism sector, as well as the difficulty in communication with residents who do not all understand what foreign tourists are communicating. According to the author, there is a need for an interactive chatbot that can assist in obtaining route information and access to the desired tourist destinations, which will facilitate foreign tourists.

Artificial intelligence must be utilized to improve the chatbot's accuracy in answering sentence selection. Several previous studies have added artificial intelligence to building chatbots. One of which is creating a chatbot related to Covid-19 [9]. Other research is also being carried out in building chatbots with artificial intelligence, namely a classification method to identify intentions rather than user input, it is called purpose classification in the chatbot system [10]. Furthermore, in the field of tourism, another research has been conducted by creating a website along with a chatbot for the city of Kanazawa [11]. The method used in this research is the Random Forest Classifier method. Random Forest Classifier is an algorithm that results from the bootstrapping aggregation of Decision Tree algorithms. This research employs this method due to its advantages over other algorithms, as it falls into Classification and Regression Tree (CART) methods, which utilize historical data to build a decision tree.

Based on the background, the author believes that an interactive chatbot capable of assisting foreign visitors in obtaining information about routes and ways to reach their desired tourist destinations will greatly facilitate them. In this study, it is expected that an interactive chatbot using the Random Forest Classifier method can optimize the accuracy level in sentence prediction performance and utilize the Telegram Messenger to structure the data more effectively while also providing social services.

II. LITERATURE REVIEW

A. Chatbot

Chatbot is a program in artificial intelligence designed to communicate directly with humans as its users. The difference between a chatbot and a natural language processing system is the algorithms' simplicity. Although many bots can interpret and respond to human input, they only interpret keywords in the input and reply with the most suitable keywords or patterns of words from pre-existing data in a database created beforehand [12]. The future of Software Engineering is expected to undergo a significant transformation with the emergence of chatbots. These chatbots will enable software practitioners to communicate and inquire about their projects using everyday language, revolutionizing how they interact with various services. At the core of every chatbot lies a Natural Language Understanding (NLU) component, which empowers the chatbot to comprehend and interpret human language inputs [13].

Initially, these computer programs (bots) were tested through the Turing test, which involved concealing their identity as machines to deceive the person conversing with them. If a user cannot identify the bot as a computer program, that chatbot is categorized as artificial intelligence.

One famous chatbot is Eliza (Dr. Eliza), developed by Joseph Weizenbaum at the Massachusetts Institute of Technology (MIT). Eliza is a pioneering chatbot known as a chat program that plays the role of a psychiatrist. Eliza simulates conversations between a psychiatrist and their patients in natural English. Eliza was created to study natural language communication between humans and machines. Eliza acts as a psychologist who can answer the patient's questions with reasonable responses or respond with further questions [12].

B. Random Forest Classifier

Random Forest is a method introduced by Breiman, a development and combination of multiple Decision Trees. While a Decision Tree represents a single classification tree, Random Forest creates multiple trees to determine its prediction results. Combining bootstrap aggregating and random feature selection in a Random Forest can reduce the overfitting problem in small training data [14]. Since Random Forest is an ensemble method of CART, it does not assume or work well in non-parametric cases.

The steps involved in the Random Forest as shown in Fig. 1 are explained as follows:

- Determining the parameters of Random Forest. The value of *mtry*, or the number of randomly selected predictor variables, is set to $\frac{p}{3}$ for regression cases, where *p* is the total number of predictor variables [14].
- Then, determine the recommended number of N_{tree} trees to use, typically 50 trees. According to Breiman [14], 50 trees provide satisfactory results for classification cases. However, $N_{tree} \geq 100$ yields lower misclassification rates [15].
- Specifying the stopping criteria default in scikit-learn Random Forest, where one means that if a subnode/child node contains only 1 sample, it will stop splitting and become a terminal node/leaf node. Therefore, once the branching stops, terminal nodes are generated as the prediction result of a single CART tree [16].
- Splitting the data into training and testing datasets. From the training data, n samples are randomly selected with replacement (bootstrap) to create a new dataset D, where I represents the bootstrap sample division of the i- tree.

- The bootstrap sampling only takes $\frac{1}{3}$ of the entire training data and the remaining $\frac{1}{3}$ is considered out-of-bag (OOB) data, which is useful for measuring the performance of regression trees [17].
- Making predictions based on constructing tree models from the new dataset D, using a combination of randomly selected m predictor variables (random feature selection).

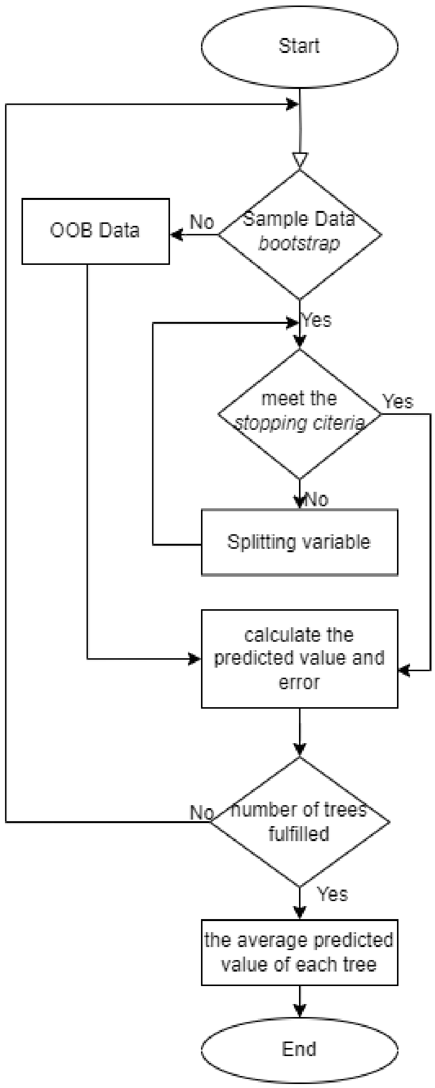

Fig. 1. Flow diagram of random forest.

III. Proposed Model

The initial stage involves data collection. Data collection for this research uses references from Google, specifically collecting frequently used conversational sentences in English. The data collection for this research is divided into two parts: training data collection and testing data collection. While collecting training data, the author used 100 conversational sentences in English. These sentences were then labeled based on their sentence classes, whether they belonged to the statement (S), question (Q), or chat (C) class. Table I provides an example of the conversational sentences used as training data, labeled based on their sentence classes.

TABLE I. Example Dataset of Sentence for Model Formation and Classification

No	Sentence	Label
1	Sorry, I don't know about the weather.	S
2	What's the weather like today?	C
3	I am fine	C
4	Where do you live?	Q
5	Are you a chatbot?	Q

After labeling the dataset consisting of 100 conversational sentences, the dataset will be processed by assigning Part-of-Speech (POS) tags to each word in the corpus. Fig. 2 displays the type of Penn treebank tagset. Once POS tags are assigned, the patterns for classifying the training sentence models will be determined. Part-of-Speech (POS) tagging, or simply tagging, is the process of assigning syntactic labels or Part-of-Speech tags to each word in the corpus. Since tags are generally applied to punctuation marks, punctuation marks such as periods, commas, etc., must be separated from the words during the tagging process. The extracted features from the data are used to build the required model by extracting Part-of-Speech (POS) tags, resulting in numerical data features. Fig. 3 provides an example of the POS tagging process in this research.

After that, the tags in each sentence will be calculated according to the predetermined tags, serving as a reference for pattern formation within the sentences. The tags used as markers in this process include a cardinal number (CD), noun singular or mass (NN), proper noun singular (NNP), proper noun plural (NNPS), noun plural (NNS), personal pronoun (PRP), verb gerund or present participle (VBG), and verb 3rd person singular present (VBZ). Table II provides an example dataset table for counting the number of Part-of-Speech (POS) tags.

Tag	Description	Example	Tag	Description	Example
CC	coord. conjunction	*and, or*	RB	adverb	*extremely*
CD	cardinal number	*one, two*	RBR	adverb, comparative	*never*
DT	determiner	*a, the*	RBS	adverb, superlative	*fastest*
EX	existential there	*there*	RP	particle	*up, off*
FW	foreign word	*noire*	SYM	symbol	*+, %*
IN	preposition or sub-conjunction	*of, in*	TO	"to"	*to*
JJ	adjective	*small*	UH	interjection	*oops, oh*
JJR	adject., comparative	*smaller*	VB	verb, base form	*fly*
JJS	adject., superlative	*smallest*	VBD	verb, past tense	*flew*
LS	list item marker	*1, one*	VBG	verb, gerund	*flying*
MD	modal	*can, could*	VBN	verb, past participle	*flown*
NN	noun, singular or mass	*dog*	VBP	verb, non-3sg pres	*fly*
NNS	noun, plural	*dogs*	VBZ	verb, 3sg pres	*flies*
NNP	proper noun, sing.	*London*	WDT	wh-determiner	*which, that*
NNPS	proper noun, plural	*Azores*	WP	wh-pronoun	*who, what*
PDT	predeterminer	*both, lot of*	WP$	possessive wh-	*whose*
POS	possessive ending	*'s*	WRB	wh-adverb	*where, how*
PRP	personal pronoun	*he, she*			

Fig. 2. Penn treebank tagset.

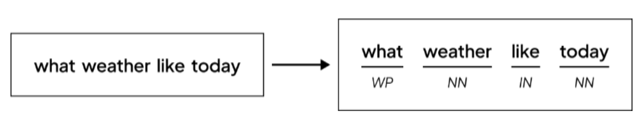

Fig. 3. Example of POS tagging.

TABLE II. EXAMPLE DATASET FOR COUNTING THE NUMBER OF PART-OF-SPEECH (POS) TAGS

No	Word Totals	CD	NN	NNP	NNPS	NNS	PRP	VBG	VBZ
1	7	0	1	1	0	0	1	0	0
2	5	0	2	0	0	0	0	0	1
3	3	0	0	0	0	0	1	0	0
4	4	0	0	0	0	0	1	0	0
5	4	0	1	0	0	0	1	0	0

After collecting the training dataset of conversational sentences and the Part-of-Speech patterns for building the classification model, the next step is to determine the dataset of conversational sentences to be used as the output sentences for the chatbot based on the formed training classification model. The dataset represents the data used in the testing process. There are 4,085 conversational sentences that will be used as the dataset in the testing process. Subsequently, this data will be labeled according to the sentence classes. The labeling is based on the criteria of whether the sentence belongs to the question (Q) or chat (C) class, similar to the labeling in the first dataset. Table III provides an example dataset table of conversational sentences that will be used in the testing process.

TABLE III. EXAMPLES OF OUTPUT SENTENCE DATASETS USED IN THE TESTING PROCESS

No	Sentence	Label
1	Hi there, how are you!?	C
2	My name is Tourist Chatbot, but you can call me Lisa	C
3	ok, thanks!	C
4	Do you like it?	Q
5	What's that?	Q

The next stage is to process the user's input in conversational sentences. The entered conversational sentences by the user will go through the processing stage, which includes preprocessing and training processes. In the preprocessing stage, a Natural Language Processing (NLP) approach is used, which involves case folding to convert all letters in the document to lowercase, tokenization to separate input words into individual tokens, stemming from finding the base form of words and producing the correct language structure, and POS tagging to assign Part-of-Speech tags or syntactic classes to each word in the corpus.

After the preprocessing process, the data will proceed to the training process, which involves training the preprocessed data using the Random Forest Classifier approach (see Fig. 4). In this data training stage, a classification model will be formed through the training using the Random Forest Classifier method. The final stage is testing, which is conducted to obtain accuracy and error values for the chatbot. Once the process is completed, the system will generate an output sentence to be sent as a response to the user's input sentence.

IV. IMPLEMENTATION

The implementation phase is the stage of applying the designed system based on the system design that has been created. The results of this system implementation are used to classify sentences in the virtual route guide chatbot using the random forest classifier method. The implementation phase begins by inputting data into the system that has been created using the Python programming language. The labeled sentence data, stored in .csv format, is inputted into the system, and the column containing the conversational sentences and sentence classes is extracted.

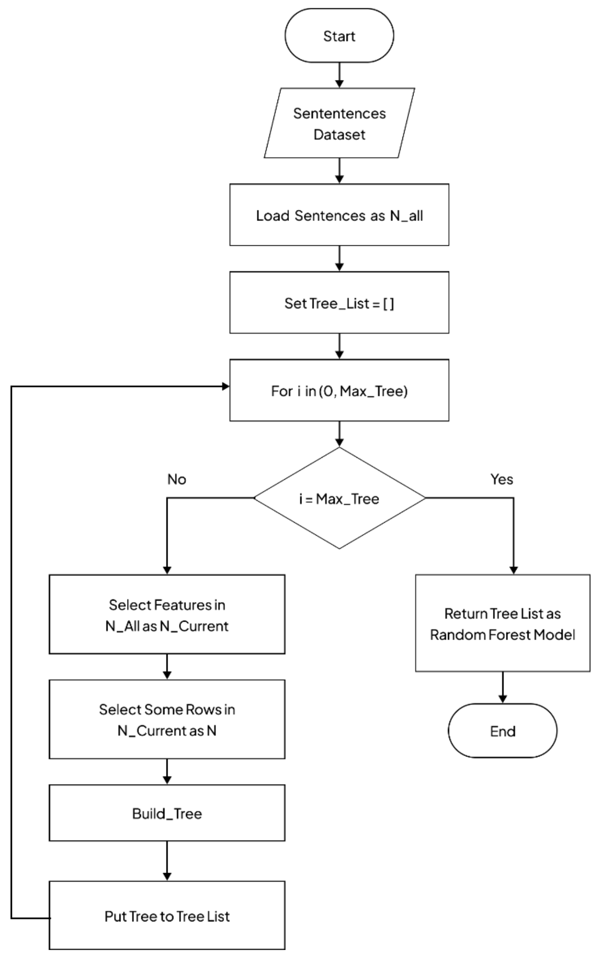

Fig. 4. Random forest classifier algorithm flowchart.

Next, the system proceeds to the initialization stage of part-of-speech patterns for sentence formation, which will be used in feature selection. Several part-of-speech patterns are used in this initialization stage, such as VerbCombos, questionTriples, statementTriples, startTuples, and endTuples. Then, feature_keys are initialized as keywords for the part-of-speech tags used.

After completing the initialization stage, the system moves to the preprocessing stage, including case folding, tokenizing, stemming, and POS tagging. The case folding process utilizes the lower() function to convert the entire sentence to lowercase to avoid case sensitivity. The tokenizing process uses the split() function to separate the sentence into individual words. The stemming process utilizes libraries from

NLTK with the WordNetLemmatizer and SnowBall algorithms. The POS tagging process assigns Part-of-Speech tags to each word in the corpus.

The next step is to create a function to count the frequency of occurrence of part-of-speech tags in each sentence. The processed data, using the function to count the frequency of part-of-speech tags, is then stored in a data frame containing the calculated frequency counts and saved in a new CSV file. This data calculates the number of part-of-speech tags in each sentence. Once the calculated frequency counts of part-of-speech tags are successfully saved, the dataset is imported and loaded into Python. Then, the class variable with the object data type is converted to the int64 data type. This is because the class variable represents a category that indicates it is categorically ordinal.

Next, the feature-selected data stored in a new file will be divided into training and testing data. The training data will be implemented into the system that has been created to build a sentence classification model using the random forest classifier. Based on the rule of thumb, the training and testing data will be split in a 75:25 ratio, with 4,185 pattern data points to be implemented in the system. This data split is done using the model_selection library's train_test_split function.

After dividing the data into training and testing sets, the next step is to train the dataset by creating a classification model using the random forest classifier. The aim is to classify data effectively and efficiently. This model is used to classify data into multiple categories or classes with high accuracy. In creating this classification model, the feature extraction function is used to improve the accuracy of the classification model by eliminating irrelevant features and retaining the most important features in the data. By selecting the most relevant and important features, the classification model can find more accurate patterns and distinguish between different classes better.

Once the training model is created, the next step is to create a Telegram account and access @BotFather. @BotFather is the official Telegram bot that allows users to create their own Telegram bot. After creating the Telegram bot, the next step is to install the python-telegram-bot library. This library is used to connect the bot to the Telegram API and send and receive user messages.

The next step is to write Python code for the bot and add message-handling functions that the bot will process. In creating the Python code, the telegram.ext module from the python-telegram-bot library is required. Then, an Updater object is created with the bot token and the bot is started using the command "updater.start_polling()". The message-handling function will be executed every time the bot receives a new message from the user. Specific logic is added to the message handling function to process user requests and provide appropriate responses based on the received message.

V. EXPERIMENTAL RESULT

In this research, there are two datasets, each consisting of 100 and 4,185 labeled conversational sentences according to the rules of Part-of-Speech (POS) tagging, based on the criteria of whether the sentences belong to the question (Q) or chat (C) class. Table IV shows the number of sentences in each label within the conversational sentence dataset used as training data for the classification model.

TABLE IV. NUMBER OF SENTENCES PER LABEL

Dataset 1		
No.	Label	Number of Sentences
1	S (Statement)	32
2	Q (Question)	43
3	C (Chat)	25
Total Number of Sentences		100
Dataset 2		
No.	Label	Number of Sentences
1	S (Statement)	1473
2	Q (Question)	1.238
3	C (Chat)	3.073
Total Number of Sentences		4.185

The data will be divided into two parts: training and testing data. The training data will be implemented into the system that has been created to build a classification model using the random forest classifier for sentences. Meanwhile, the testing data is used to evaluate the system's performance. The conversational sentence dataset will be divided into testing data and training data with a ratio of 75:25 based on the rule of thumb. In addition, experiments are conducted by varying the number of trees (n_estimators), tree depth (max_depth), and the minimum number of samples required to split a node (min_sample_split). The number of trees tested is 100, 200, and 500, the depth of the trees tested is 5, 10, and 20, and the minimum samples required to split a node tested are 2, 5, and 10.

From Table V, the highest accuracy value is obtained using 200 trees, with a maximum tree depth of 20 and a minimum sample split of 5. Using these parameters yields an accuracy value of 95.88%, precision of 96.29%, recall of 96.03%, and f-measure of 96.16%. On the other hand, the lowest accuracy value is obtained from using 500 trees, with a maximum tree depth of 5 and a minimum sample split of 10. Using these parameters yields an accuracy value of 94.43%, precision of 95.29%, recall of 94.61%, and f-measure of 94.95%. These accuracy results are also depicted in the visualization graph, as shown in Fig. 5.

TABLE V. SHOWS THE TESTING RESULTS USING 100, 200, AND 500 TREES WITH A COMBINATION OF 20 MINIMUM SAMPLES AND 10 MINIMUM SAMPLES REQUIRED TO SPLIT A NODE

Tree	**Deep *Max. Tree***	***Min. Sample***	***Accuracy***	***Precision***	***Recall***	***F-measure***
100	5	2	94,58%	95,30%	94,80%	95,05%
		5	94,73%	95,40%	94,95%	95,17%
		10	94,66%	95,35%	94,87%	95,11%
	10	2	95,42%	95,94%	95,61%	95,78%
		5	95,34%	95,89%	95,53%	95,71%
		10	95,11%	95,70%	95,32%	95,51%
	20	2	95,50%	95,83%	95,73%	95,78%
		5	95,73%	96,12%	95,91%	96,02%
		10	95,50%	96,00%	95,65%	95,82%
200	5	2	94,66%	95,45%	94,83%	95,14%
		5	94,73%	95,50%	94,91%	95,21%
		10	94,58%	95,40%	94,76%	95,08%
	10	2	95,57%	96,08%	95,75%	95,92%
		5	95,50%	96,07%	95,65%	95,86%
		10	95,42%	96,01%	95,57%	95,79%
	20	2	95,42%	95,74%	95,71%	95,72%
		5	95,88%	96,29%	96,03%	96,16%
		10	95,57%	96,05%	95,73%	95,89%
500	5	2	94,50%	95,35%	94,69%	95,02%
		5	94,50%	95,35%	94,69%	95,02%
		10	94,43%	95,29%	94,61%	94,95%
	10	2	95,50%	96,00%	95,69%	95,84%
		5	95,42%	95,98%	95,59%	95,78%
		10	95,42%	95,98%	95,59%	95,78%
	20	2	95,50%	95,76%	95,82%	95,79%
		5	95,80%	96,24%	95,95%	96,09%
		10	95,57%	96,05%	95,73%	95,89%

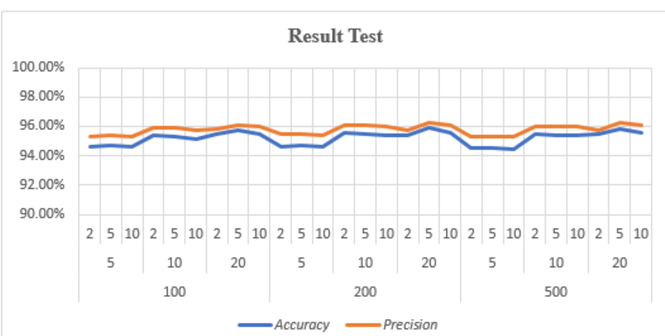

Fig. 5. Visualization of test results in conversational sentences.

VI. CONCLUSION

The result of this research is an interactive chatbot created using the Random Forest Classifier method, which can optimize the level of accuracy in sentence prediction performance and utilizes Telegram Messenger to structure the data more effectively. According to the test results in the previous section, this study obtained the highest accuracy value using a tree quantity of 200, a maximum tree depth of 20, and a minimum sample split of 5. Using these quantities resulted in an accuracy of 95.88%, precision of 96.29%, recall of 96.03%, and f-measure of 96.16%. This proves that the use of random forest classification affects the sentence classification results. The reason is the existence of feature selection that can reduce the features used.

ACKNOWLEDGMENT

We express our deepest gratitude to all members of the Informatic Engineering, Faculty of Science and Technology, UIN Maulana Malik Ibrahim, Malang.

REFERENCES

[1] Ismayanti, Introduction to Tourism. Jakarta: Grasindo, 2010.

[2] Tugu Malang, "6 million tourists are visiting Malang, the number exceeds the 2022 target," Tugu Malang, 2022.

[3] Communication and Informatics Office of Malang Regency, KABUPATEN MALANG SATU DATA. 2022.

[4] Statistics of Malang Municipality, “Number of Domestic Tourists in Malang City (People) in 2020-2022,” https://malangkota.bps.go.id/indicator/16/157/1/jumlah-wisatawan-domestik-di-kota-malang.html, 2022.

[5] Statistics Batu City, “Number of Visitors to Tourist Attractions and Tourism Souvenirs According to Tourist Attractions in Batu City in 2021.” [Online]. Available: https://batukota.bps.go.id/statictable/2022/04/11/1383/jumlah-pengunjung-objek-wisata-dan-wisata-oleh-oleh-menurut-tempat-wisata-di-kota-batu-2021.html. [Accessed: 26-Apr-2023].

[6] M. Aziz and M. Aman, “Decision Support System For Selection Of Expertise Using Analytical Hierarchy Process Method,” 2019.

[7] D. Rinova, “Proceedings Sustainable Development Goals (SDGs) Conference International Science Consortium for Indonesian Sustainability (ISCIS) Analysis of Tourist Attraction and Service Quality on Tourist Satisfaction.”

[8] M. A. Cholik, “THE DEVELOPMENT OF TOURISM INDUSTRY IN INDONESIA : CURRENT PROBLEMS AND CHALLENGES,” Eur. J. Res. Reflect. Manag. Sci., vol. 5, no. 1, 2017.

[9] W. Astuti, D. P. I. Putri, A. P. Wibawa, Y. Salim, Purnawansyah, and A. Ghosh, “Predicting Frequently Asked Questions (FAQs) on the COVID-19 Chatbot using the DIET Classifier,” 3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIConCIT 2021, pp. 25–29, 2021, doi: 10.1109/EIConCIT50028.2021.9431913.

[10] M. Y. H. Setyawan, R. M. Awangga, and S. R. Efendi, “Comparison Of Multinomial Naive Bayes Algorithm And Logistic Regression For Intent Classification In Chatbot,” Proc. 2018 Int. Conf. Appl. Eng. ICAE 2018, pp. 1–5, 2018, doi: 10.1109/INCAE.2018.8579372.

[11] D. Suzuki, K. Nunotani, K. Fukusato, and M. S. Tanaka, “A Study of Tourism Proposal System Using AI,” 2020 IEEE 9th Glob. Conf. Consum. Electron. GCCE 2020, pp. 634–635, 2020, doi: 10.1109/GCCE50665.2020.9292070.

[12] J. Weiznbaum, “ELIZA - A Computer Program for the Study of Natural LanguageCommunication Between ManAnd Machine,” Commun. ACM, vol. 9, no. 1, pp. 36–45, 1996.

[13] A. Abdellatif, K. Badran, D. E. Costa, and E. Shihab, “A Comparison of Natural Language Understanding Platforms for Chatbots in Software Engineering,” IEEE Trans. Softw. Eng., vol. 48, no. 8, pp. 3087–3102, 2022, doi: 10.1109/TSE.2021.3078384.

[14] L. Breiman, J. Friedman, C. J. Stone, and R. . Olshen, Classification and Regression Trees, 1st ed. Boca Raton : Taylor & Francis Group, 1984.

[15] C. D. Sutton, “Classification and Regression Trees, Bagging, and Boosting,” Handbook of Statistics, vol. 24. Elsevier, pp. 303–329, 2005, doi: 10.1016/S0169-7161(04)24011-1.

[16] F. Widmaier, A. Zell, A. Schilling, and J. Bohg, “Robot Arm Tracking with Random Decision Forests.”

[17] S. Liu et al., “Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization,” Comput. Electron. Agric., vol. 95, pp. 82–91, 2013, doi: 10.1016/j.compag.2013.03.009.