MyBotS Prototype on Social Media Discord with NLP

Imam Al Maksur  Muhammad Muhajir *

Department of Statistics, Universitas Islam Indonesia, Sleman, Yogyakarta, Indonesia.

*Corresponding author: 16611064@students.uii.ac.id , mmuhajir@uii.ac.id *

*ORCID ID: https://orcid.org/0000-0003-0512-0899, https://orcid.org/0000-0001-7576-2630*

Received 21/12/2020, Accepted 15/3/2021, Published 30/3/2021

This work is licensed under a Creative Commons Attribution 4.0 International License.

Abstract:
The continuous growth in technology and technological devices has led to the development of machines to help ease various human-related activities. For instance, irrespective of the importance of information on the Steam platform, buyers or players still get little information related to the application. This is not encouraging despite the importance of information in this current globalization era. Therefore, it is necessary to develop an attractive and interactive application that allows users to ask questions and get answers, such as a chatbot, which can be implemented on Discord social media. Artificial Intelligence is a technique that allows machines to think and be able to make their own decisions. This research showed that the discord chatbot prototype provides various services based on the results of classification testing using the SVM method with three kernels, namely Linear, Polynomial, and RBF. The test data and accuracy values prediction are the largest Linear Kernel SVM with accuracy and error prediction values of 94% and 6%.

Key words: Artificial Intelligence, Chatbot, Discord, NLP, SVM

Introduction:
All over the world, technology and technological devices tend to continuously grow with the development of various machines created to help ease various human-related activities. Various studies have shown that buyers or players on the Steam platform still get little information related to games. Irrespective of the importance of information in the current globalization era, loyal Steam players are not provided with the necessary amount. Therefore, based on this problem, it is necessary to develop an application that allows users to ask questions and get adequate replies. For instance, questions regarding game prices, ratings, new game releases dates, etc. The application designed to overcome this issue is the Chatbot, which is implemented on Discord social media and provides adequate assistance to online game players. This device can be installed on Android, Mac, PC, etc. Discord is a new application that is simple, practical, attractive, easy to use and accessible from various gadgets (1). Similar to Whatsapp, Telegram and Skype, this application provides voice chat assistance to gamers and streamers that allows for Voice Over Internet Protocol (VOIP) and messaging among users (2). The application also allows users to contribute through various listening sources. Therefore, by using this application, it is believed that students are likely to shed off the traditional role of being passive receptors (3).

The operative method of the Chatbot was carried out in the research design and manufacture of its prototype using an Artificial Intelligence (AI), which is a technology that makes it possible for computers to think and make their own choices. According to Martinez-Plumed et al. (2018), it is not difficult to have computers that are capable of thinking and behaving like humans using artificial intelligence (4). The Chatbot application uses the Natural Language Processing (NLP) programming language which is a field of computer science, AI, and computational linguistics that bridges the gap between clinical human language and computer systems (5).

A 2020 study carried out by Zahour et al. aimed to prepare chatbots based on the theory of John Holland and the RIASEC questionnaire in the field of education and professional guidance, to assess the superiority of the personality styles of undergraduate and postgraduate students seeking to join the labor market (6).

Furthermore, bot architecture research applied to the Discord application was carried out by Septiansyah et al. in 2018 (7). The study was
Artificial intelligence markup language (AIML)

Artificial Intelligence Markup Language (AIML) aims to facilitate the modeling of communication dialogue (18). AIML defines data objects that are used to model the input and output patterns of a conversation. Furthermore, it represents the knowledge used by Chatbots based on the development of ALICE bot technology (19).

Text mining

Text mining is mines data typically derived from documents in the form of text (20) and it is shown in Fig. 1. The preprocessing stage is generally needed in text mining on documents (21).

Figure 1. Text Mining Process

Case folding is the process of changing capital letters in a document to lowercase which includes the letters "a" to "z" and eliminating other characters considered delimiters (22). The case folding process also helps to change the capital letters in the MyBotS dataset to lowercase.

Tokenizing is the process of truncating input strings in text based on the composed words (22). Therefore, this process leads to the generation of words that have been separated in a sentence MyBotS dataset.

Filtering process takes important words or those required for analysis using a stoplist/stopword or wordlist algorithm (22).

Stemming is the process of changing affix into basic words, and this is generally carried out in English text (22). This process is important in MyBots application because it uses English.

Word weighting TF-IDF

Term frequency-inverse document frequency (TF-IDF) is a numerical statistic that tells users the importance of a word in a list or corpus document in various processes for information recovery and text mining (23–27). It is also used as a weighting factor with different types of algorithms used to construct an automated description (28). Term Frequency-Inverse Document (TF-IDF) is the most widely used extractive text description (29).

The type of formula to be used for calculating term frequency (TF) is pure TF (raw TF) because it shows how often a word appears in a particular document (30). Meanwhile, IDF explains the importance of a word in a document which is formulated as follows:

\[ IDF(A) = \log \frac{D}{df_A} \] (1)

When the value is similar, zero results for the IDF calculation are obtained. Therefore, a value of 1 (+1) is added to the IDF formula as follows:
\[ IDF(A) = \log \frac{n}{df_A} + 1 \quad (2) \]
\[ W(A) = TF(A) \times IDF(A) \quad (3) \]

**Support vector machine (SVM)**

Different methods and algorithms of data mining are applied to the development of data classifiers. The most popular tools for data mining include artificial neural networks, decision trees, and SVM algorithm (Support Vector Machine Algorithm) (31). SVM is used to make classification and regression predictions (32).

Suppose there are \( n \) variables, and \( \{ x \in \mathbb{R}^n, \ i = 1, 2, \ldots, m \} \), the input vector is set by building a hyperplane, SVM constructs and a regression model, i.e.,

\[ f(x) = \sum_{i=1}^{n}(a_i^* - a_i)k(x_i, x) + b \quad (4) \]

where \( a^* \) and \( a_i \) are the Lagrange multipliers, \( b \) is offset quantity and the kernel function

\[ k(x, x) = \varphi(x_i)\varphi(x) \quad (5) \]

Where \( y_{n} = 1 / -1 \) is a constant which denotes the class of each point \( x_n \) with \( n \) = number of samples. Each \( x_n \) is \( p \)-dimensional real vector, with scaling used to safeguard variables with greater variety. This training data can be viewed by dividing (or separating) the hyperplane, which requires

\[ w \cdot x + b = 0 \quad (6) \]

Where the scalar is \( b \) and the \( p \)-dimensional is \( w \), which refers to the perpendicular hyperplane. The margin is increased by adding the offset parameter \( b \), with the hyperplane mandated to move through the origin without \( b \), thereby restricting the solution. The stages or steps of the classification evaluation analysis in this study are shown in the flowchart in Fig. 2.

**Methods:**

**The proposed method**

This section introduced an algorithm scheme for designing and making prototypes in the field of computer science, namely AI. The prototypes which were implemented in the form of a chatbot on the Discord social media application used NLP as the newest programming language to make and apply SVM for MyBotS Classification accuracy analysis. The flow process of designing an algorithm and making prototypes of MyBotS is shown in Fig. 3.
Research method:
Chatbot system was created for game players to obtained relevant information regarding game sales on the Steam platform using the discord social media. It uses NLP as a language-based social media discord program and the STEAM online platform to obtain game sales datasets within 24 hours. Therefore, based on the objectives of the system, the following features need to be included in the Chatbot:

a. Functional requirements: 
1) Steam Store information, 2) Steam Event, 3) Clan Event, 4) Music Bot, and 5) a relaxing game features of Ball8 Game Bot.

b. Non-functional requirements: 1) This chatbot system can be accessed on Discord social media, and 2) has a user-friendly interface.

c. Software requirements: 1) Visual Studio Code, 2) Sniping Tools, 3) Java Script nodes, 4) Discord, 5) Steam, 6) Ffmpeg, and 7) Browser

d. Hardware requirements: 1) CPU: Intel® Core™ i5-7300HQ CPU @ 2.50GHz, 2) GPU: NVIDIA GeForce GTX 1050 4GB, 3) RAM: Team T-Force Vulcan SODIMM DDR4 16Gb, 4) Hard drive: 2 TB, and 5) Laptop: LEGION Y520

In designing the system, a database is needed to store goods and user. The database used to carry out this research is json which saves directly to a private folder.

The steps or stages in this research are shown in the flowchart in Fig. 4.

Results and Analysis: Problem and implementation of the Chatbot System

Natural Language Processing (NLP) is a method used to implement Chatbot in solving STEAM platform information problems. Chatbot application also needs information retrieval (IR) to obtain information when questioned. This chatbot system was designed using the graph master pattern matching method, which serves as an answer search algorithm. In the Analysis phase, the Chatbot built provides information regarding STEAM STORE to users on Discord social media. The context diagram comprising of the highest DFD level representing the system input or output provides a summary of the entire system. Chatbot's background diagram is shown in Fig. 5.

Figure 4. Research Flowchart

Figure 5. Context Diagram Chatbot

Figure 5, shows that the context diagram has two entity, namely admin and users. Admin needs to first log into the developer Discord system in order to access the Chatbot script. They are tasked with managing data, maintaining and repairing Chatbot. Admin can also manage data script, and responsible for various activities, such as Create, Read, update, delete (CRUD) and search. They have the ability to carry out a chatbot test on MyBotS test channel, which is one of the channels that can only be accessed by the Admin or Owner. Meanwhile, users of Discord can only have a conversation or
chat by entering sentences or questions on the features of the MyBotS Chatbot.

**Natural Language Processing**

The NLP system consists of 3 parts, namely input, knowledge base and output. After the Chatbot is invited into the discord server, the user greets it with the word, hello to start a conversation. This is followed by a reply prompting users to input further details to be processed, which is searched to determine matching knowledge base on what is contained in the chatbot script. When the appropriate data is found, bot proceeds to provide appropriate answers to the user input. After the answer is found, it is displayed on the output channel. The flow diagram of the chatbot system is shown in Fig. 6.

**Knowledge Base Chatbot**

In the chatbot system analysis is given, which is used as a knowledge base, which is tabulated, as shown in Table 1.

**Table 1. Chatbot knowledge base**

<table>
<thead>
<tr>
<th>No</th>
<th>Pattern</th>
<th>Template</th>
</tr>
</thead>
</table>
| 1 | switch (args[0]) {  
  case "hallo":  
  break;} | "hallo! My Name MyBotS will help you about information of STEAM STORE and another future, please check '!J help' to see anymore commands on this bot, My Owner Imam Al Maksur");} |
| 2 | (msg01.content = '!JSteam Store'}) | Steam Store, Top Seller Games, the best collection of game sales on the Steam Platform. |
| 3 | ('message', msg =>{ (msg.content '!J Steam Game "nama game")} | Game name, game description, current game price, Publisher, Developer, Genre, Release Date, and user review |
| 4 | ("message", asyncmsg =>  
  {msg.content === '!J Event) | Information related to the event that will be held next |
| 5 | ("message", asyncmsg =>  
  {msg.content === '!J Date) | To determine the date and time of upcoming events. |
| 6 | case 'play':  
  function play(connection, msg){  
  {connection.disconnect();} | you need to provide a link | you must be in the channel to use the bot! MyBotS akan memainkan musik req di Voice Channels |
| 7 | case 'skip':  
  var server = servers[msg.guild.id];  
  server.dispatcher.end();  
  break; | skip the song! |
| 8 | case 'stop':  
  var server = servers[msg.guild.id];  
  (msg.guild.voiceConnection){break; | End the queue leaving the voice channel! |
| 9 | ("message", msg => {constmsgToLowerCase = msg.content.toLowerCase();  
  constgameArray = ['rock', 'paper', 'scissors']; | "It's a draw! ṭ", "The computer wins! ṭ", "You win! ṭ";  
  "You've chosen rock!", "You've chosen paper!", "You've chosen scissors!";  
  Pc has chosen: |

**Chatbot Prototype Response**

The chatbot program that has been built produces a simple response on information about the Steam Store, Steam Game, Event, Music bot and paper rock scissors game, as shown in Table 2.
Table 2. Chatbot Prototype Response

Opening

List command MyBotS

Steam Store
Steam Game

![Image of Steam Game details]

Steam Event

![Image of Steam Event details]

Event Date

![Image of Event Date details]

Music Bot

![Image of Music Bot details]
Support Vector Machine

Prior to the classification process, data is first separated into training and testing. The training and testing data used were 80% and 20% of total data, i.e. 67 and 17, which were classified using the Vector Machine kernel algorithm. Kernels are Linear, Poly and RBF, therefore, before classifying the data, TF-IDF words were preprocessed and weighed. The classification process involves training data to construct a model, which is used for testing. Classification of Support Vector Machine uses 3 kernels, namely Linear, Poly, and RBF. The results of the SVM classification comparison are summarized in Table 3.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Accuracy</th>
<th>Error</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>100%</td>
<td>0%</td>
<td>Training</td>
</tr>
<tr>
<td>Linear</td>
<td>94%</td>
<td>6%</td>
<td>Testing</td>
</tr>
<tr>
<td>Polynomial</td>
<td>86%</td>
<td>14%</td>
<td>Training</td>
</tr>
<tr>
<td>Polynomial</td>
<td>64.7%</td>
<td>35.3%</td>
<td>Testing</td>
</tr>
<tr>
<td>RBF</td>
<td>97%</td>
<td>3%</td>
<td>Training</td>
</tr>
<tr>
<td>RBF</td>
<td>82.3%</td>
<td>17.7%</td>
<td>Testing</td>
</tr>
</tbody>
</table>

Table 3 shows that comparing SVM classifications using three different kernels shows that the Linear kernel has a higher accuracy of 100% and 94% for training and testing data. Furthermore, to ensure that the Linear kernel has better accuracy than the Polynomial and RBF kernels for this study Discord-based social media chat Discord was developed by providing various service categories, such as Steam Shop, Steam Case, Music Bot and Ball8 Game both categories.

Predict categories on new question dataset

After determining the results of the SVM classification model with a linear kernel, the model is evaluated using new question data, which is not included in the training data.
Figure 7 shows that the evaluation of prediction is in a very good category with 9 more questions that are not misclassified. However, despite this classification, the Chatbot still needs to be trained.

**Descriptive Statistics**

Figure 8a shows that the most frequently asked questions are in the Steam Store category, with a percentage of 68%. The Music bot category follows this at 18%. Meanwhile, the least asked questions were in the Game Bot and Steam Event category with a percentage of 7%.

Overall, 84 questions were asked by Discord users. Figure 8b showed that the Chatbot answered the questions asked by the user correctly at a percentage of 93% while the wrong answers were 7%.

**Conclusion:**

In conclusion, discord-based social media chat has been built by providing various service categories, such as Steam Store, Steam Event, Music bot and Ball8 Game bot. Furthermore, based on the results of the Support Vector Machine (SVM) classification analysis with three kernels, namely Linear, Polynomial, and RBF kernels with testing data, the accuracy value of the prediction category is greatest using SVM with a Linear kernel, of 94% and prediction error of 6%. The chatbot program system provides answer responses with an accuracy value of 100% and an error of 0% from 10 new questions.

**Acknowledgment:**

The authors are grateful to the Department of the Statistics Islamic University of Indonesia for their support throughout the research process.

**Authors' declaration:**

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are mine ours. Besides, the Figures and images, which are not mine ours, have been given the permission for republication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in Universitas Islam Indonesia.

**References:**

2. Lacher L, Biehl C. Using Discord to Understand and Moderate Collaboration and Teamwork. Proc 49th


النموذج المبدئي ل MyBotS على وسائل التواصل الاجتماعي Discord

محمد مهاجر
قسم الإحصاء ، الجامعة الإسلامية إندونيسيا ، سليمان ، يوجياكارتا ، إندونيسيا

الخلاصة:
أدى النمو المستمر في التكنولوجيا والأجهزة التكنولوجية إلى تطوير الآلات للمساعدة في تسهيل الأنشطة المتعلقة بالبشر. على سبيل المثال ، بغض النظر عن أهمية المعلومات على منصة Steam ، لا يزال المشترون أو اللاعبون يحصلون على القليل من المعلومات المتعلقة بالتطبيق. هذا غير مشجع على الرغم من أهمية المعلومات في عصر العولمة الحالي. لذلك ، من الضروري تطوير تطبيق جاذب وتفاعل يسمح للمستخدمين بطرح الأسئلة والحصول على إجابات ، مثل chatbot ، والذي يمكن تنفيذه على وسائل التواصل الاجتماعي هو تقنية تسمح للآلات بالتفكير والقدرة على اتخاذ قراراتها الخاصة. أظهر هذا البحث أن نموذج chatbot الخاص ب Discord يوفر خدمات متنوعة بناءً على نتائج اختبار التصنيف باستخدام طريقة SVM بالثلاث نوى ، و Linear Kernel SVM. تم استخدام الاختبار وتنبؤ قيم الدقة أكبر 94٪ و6٪.

الكلمات المفتاحية: الذكاء الاصطناعي ، روبوت الدردشة ، Discord ، البرمجة اللغوية العصبية