

ABSTRAK

Pengembangan sistem *chatbot* menggunakan metode *pattern matching* memiliki beberapa kekurangan, seperti respon yang dihasilkan oleh *chatbot* sangat berpengaruh terhadap pola-pola kalimat/pertanyaan yang sudah didaftarkan, sehingga pola kalimat yang didaftarkan harus beragam. Oleh karena itu, penggunaan pendekatan *Natural Language Processing* (NLP) pada *chatbot* sangat membantu dalam proses memahami informasi yang diberikan oleh pengguna bahasa alami. Salah satu cara untuk memahami informasi tersebut adalah dengan melakukan klasifikasi *intent* atau klasifikasi pesan. Proses klasifikasi pada umumnya dilakukan menggunakan algoritma *machine learning*, seperti *Support Vector Machine* (SVM). Algoritma SVM memiliki performa yang sangat baik dalam hal klasifikasi dan regresi. Namun, hal tersebut hanya berlaku dalam melakukan klasifikasi biner. Dalam pengembangan *chatbot*, diperlukan sebuah algoritma yang dapat melakukan klasifikasi pada data yang memiliki lebih dari dua kelas (nonbiner). Oleh sebab itu, SVM kemudian dikembangkan dengan menggunakan pendekatan *multiclass* dan menghasilkan algoritma *Multiclass Support Vector Machine* (*Multiclass SVM*).

Pada penelitian ini, algoritma tersebut selanjutnya akan digunakan untuk melakukan klasifikasi teks pesan yang dikirimkan oleh pengguna pada *chatbot* pelayanan informasi berdasarkan kategori informasi yang ada di jurusan Informatika UPN “Veteran” Yogyakarta. Data yang digunakan pada penelitian ini terdiri dari 950 data yang terbagi ke dalam 5 kategori, yaitu administratif, dokumen, jadwal, kegiatan, dan sapaan. Data tersebut kemudian akan dilakukan *text preprocessing*, ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan selanjutnya akan dilakukan proses *training* model menggunakan algoritma *Multiclass SVM*. Keseluruhan data tersebut, selanjutnya akan dibagi menjadi data latih dan data uji dengan perbandingan 75% data latih dan 25% data uji.

Penggunaan pendekatan NLP dan algoritma *Multiclass SVM* sebagai algoritma klasifikasi pada sistem *chatbot* pelayanan informasi jurusan Informatika UPN “Veteran” Yogyakarta memberikan hasil yang baik. Berdasarkan hasil pengujian yang dilakukan, algoritma *Multiclass SVM* mendapatkan nilai akurasi sebesar 87%, nilai presisi sebesar 89% dan nilai *recall* sebesar 87%. Nilai-nilai tersebut didapatkan dari model *Multiclass SVM* yang menggunakan kernel RBF dengan parameter nilai $\gamma = 0.01$, parameter nilai $C = 100$, dan metode pendekatan *multiclass One Versus All* (OVA).

Kata kunci: *Chatbot*, NLP, *Multiclass SVM*

ABSTRACT

The development of a chatbot system using the pattern matching method has several shortcomings, such as the response generated by the chatbot greatly affects the sentence patterns/questions that have been registered, so that the sentence patterns registered must vary. Therefore, the use of the Natural Language Processing (NLP) approach in chatbots is very helpful in the process of understanding the information provided by natural language users. One way to understand this information is to perform an intent classification or message classification. The classification process is generally carried out using machine learning algorithms, such as Support Vector Machine (SVM). The SVM algorithm has a very good performance in terms of classification and regression. However, this only applies to binary classification. In the development of chatbots, we need an algorithm that can classify data that has more than two classes (non-binary). Therefore, SVM was then developed using a multiclass approach and resulted in the Multiclass Support Vector Machine (Multiclass SVM) algorithm.

In this study, the algorithm will then be used to classify the text messages sent by users on the information service chatbot based on the information category in the Informatics department of UPN "Veteran" Yogyakarta. The data used in this study consisted of 950 data divided into 5 categories, namely administrative, documents, schedules, activities, and greetings. The data will then be processed by text preprocessing process, feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF) and model training process using the Multiclass SVM algorithm. The whole data will then be divided into training data and test data with a comparison of 75% training data and 25% test data.

The use of the NLP approach and the Multiclass SVM algorithm as a classification algorithm in the information service chatbot system of the Informatics Department of UPN "Veteran" Yogyakarta gave good results. Based on the results of the tests carried out, the Multiclass SVM algorithm obtained an accuracy value of 87%, a precision value of 89% and a recall value of 87%. These values are obtained from the Multiclass SVM model that uses the RBF kernel with the parameter value $\gamma=0.01$, the parameter value $C=100$, and the One Versus All (OVA) multiclass approach method.

Keywords: *Chatbot, NLP, Multiclass SVM*