

## INTISARI

### KLASIFIKASI PERTANYAAN PADA SISTEM TANYA JAWAB BERBAHASA INDONESIA MENGGUNAKAN *NAÏVE BAYES* *CLASSIFIER*

Oleh

Yohanes Lauda Adiswara  
14/365993/PA/16186

Penelitian tentang sistem tanya jawab hingga saat ini banyak didominasi oleh Bahasa Inggris. Untuk Bahasa Indonesia, bidang sistem tanya jawab masih jarang dieksplor meskipun Bahasa Indonesia adalah bahasa resmi yang digunakan oleh lebih dari 250 juta orang. Penelitian mengenai sistem tanya jawab berbahasa Indonesia dimulai sejak tahun 2005 dengan rata-rata 2 penelitian per tahun hingga tahun 2017. Penelitian-penelitian tersebut mengambil berbagai pendekatan seperti analisis semantik, *case based reasoning*, dan *machine learning*.

Sistem tanya jawab berbahasa Indonesia pada penelitian ini menggunakan pendekatan *machine learning*. Metode pendekatan *machine learning* yang diambil adalah *naïve bayes classifier* (NBC). Pendekatan ini diambil karena memiliki nilai akurasi yang tinggi dalam klasifikasi teks merujuk pada penelitian sebelumnya. Data yang digunakan dalam penelitian ini adalah pertanyaan yang dikumpulkan melalui *platform* Google Form dengan *domain* jaringan komputer. Data dibagi menjadi data latih dan data uji dengan proporsi 50% sebagai data latih dan 50% sebagai data uji.

Hasil pengujian menunjukkan bahwa untuk evaluasi pelatihan sistem dengan *stemming* menghasilkan akurasi jawaban sebesar 100% dan tanpa *stemming* menghasilkan akurasi jawaban sebesar 96,43%. Kemudian untuk hasil pengujian sistem dengan pertanyaan baru menunjukkan akurasi jawaban sebesar 92,86 persen saat menggunakan *stemming* dan menghasilkan akurasi sebesar 85,71% ketika tidak menggunakan *stemming*.

**Kata Kunci:** *Sistem Tanya Jawab, Naïve Bayes Classifier, Klasifikasi Pertanyaan*

## ABSTRACT

### QUESTION CLASSIFICATION ON INDONESIAN QUESTION ANSWERING SYSTEM USING NAÏVE BAYES CLASSIFIER

By

Yohanes Lauda Adiswara  
14/365993/PA/16186

Research on question answering system has been done mainly in English. For Indonesian, it is rarely explored although Indonesian is official language which used by more than 250 million people. Research on Indonesian question answering system began since 2005 with approximately 2 research every year. That research using many approach such as semantic analysis, case based reasoning, and machine learning.

Indonesian question answering system which developed on this research is using machine learning approach. Machine learning method which used is naïve bayes classifier. This approachment have good accuracy based on previous research on text classification. Data which used on this research are gathered from Google Form platform with computer network domain. Data is divided into train data and test data with proportion 50% of train data and 50% of test data.

Evaluation result indicate that system's training work well with 100% accuracy for system by stemming and 96,43% accuracy for system without stemming. Then evaluation result for new question yield 92,86% accuracy for system by stemming and 85,71% for system without stemming.

**Keyword:** *Question Answering System, Naïve Bayes Classifier, Question Classification*