Abstrak

Setiap manusia memiliki banyak pertanyaan didalam kehidupannya. Terkadang pertanyaan tersebut dapat diselesaikan sendiri, namun untuk pertanyaan yang mereka tidak dapat menjawabnya, manusia membutuhkan bantuan sesamanya untuk menjawab pertanyaan tersebut. Karena itu muncul berbagai kelompok tanya jawab (*Community Question Answering*) yang dapat membantu menjawab pertanyaan tersebut. Maraknya penggunaan internet saat ini membuat banyak forum tanya-jawab (*Community Question Answering Online*) bermunculan. Bahkan forum tanya jawab yang muncul bukan hanya untuk masalah akademik, tetapi tentang kehidupan disuatu negara seperti *QatarLiving Forum*. Namun, tidak semua jawaban yang diberikan sesuai dengan pertanyaan yang diajukan. Membutuhkan waktu yang lama untuk menentukan jawaban yang sesuai dengan pertanyaan yang diajukan. Untuk itu, dibutuhkan suatu sistem yang dapat memberikan peringkat pada jawaban untuk membantu merangking jawaban yang sesuai dengan pertanyaan.

Penelitian ini mengacu pada SemEval 2016: *Semantic Evaluation Exercise Task 3 Subtask A* mengenai kemiripan antara kalimat pertanyaan dan jawaban (*Question-Comment Similarity*). Adapun tahapan yang dilakukan pada penelitian ini yaitu, dimulai dari *preprocessing* dataset berdasarkan *SemEval 2016* question answering similarity, melakukan ekstraksi fitur untuk membantu proses klasifikasi dengan menggunakan *lexical similarity feature*, *semantic similarity feature*, *non textual feature grup* dan *Heuristic*. *Classifier* yang digunakan pada penelitian ini adalah *Logistic Regression* dan *Support Vector Machine* (SVM). Nilai dari hasil klasifikasi digunakan untuk memberikan peringkat jawaban yang paling sesuai dengan pertanyaan. Pengukuran performansi pada sistem dilakukan dengan menggunakan perhitungan *Mean Average Precision*(MAP).

Hasil penelitian yang diperoleh menunjukkan pengaruh fitur terhadap kedekatan antara pertanyaan dan jawaban. *Lexical similarity feature* terutama *sub feature Cosine similarity* dan LCS menunjukkan semakin tinggi nilai *feature* pada jawaban semakin jawaban tersebut mendekati pertanyaan yang diajukan. Sedangkan nilai *semantic similarity* menggunakan *Wu Palmer Algorithm*, persebaran nilai antara kelasnya lebih merata, sehingga cukup sulit untuk membedakan ciri setiap kelasnya. *Non Textual Feature Group* membantu dalam melakukan klasifikasi jawaban dan meningkatkan akurasi sebanyak 4%. Jika dibandingkan dengan peserta SemEval 2016 Task 3, penelitian ini memiliki nilai MAP diurutan ke 8 dari 13 peserta.

Kata Kunci: *Community Question Answering, Question Answering System, similarity measure, classifier,* peringkat jawaban CQA, MAP.