ABSTRAK

Community question answering(CQA), seperti Qatar Living Forum telah menjadi salah satu tempat bagi pengguna internet untuk mendapatkan informasi. Satu pertanyaan pada CQA dapat memiliki banyak jawaban, namun pengguna harus memilih jawaban yang paling sesuai dengan pertanyaannya secara manual yang membutuhkan waktu cukup lama.Permasalahan ini dapat diatasi dengan membangun sistem pemeringkatan jawaban dari list jawaban yang telah ada dimana jawaban yang sesuai dengan pertanyaan harus diletakan diatas jawaban yang tidak sesuai, Sehingga dapat membantu pengguna mendapatkan jawaban yang paling tepat secara cepat.

Proses pertama yang akan dilakukan dalam penelitian yaitu teks preprocessing yang meliputi *tag removal,case folding tokenizing, filtering,* dan *stemming*. Setelah data teks menjadi lebih teratur dilakukan ekstraksi fitur, dimana fiturnya adalah tekstual fitur dan pemodelan topik .Tekstual fitur adalah mengidentifikasi ciri ciri sebuah jawaban dengan melihat elemen-elemen teksnya seperti melihat apakah sebuah jawaban mengandung tanda tanya(?), emotikon,link atau kata-kata tertentu.Pemodelan topik merupakan pemodelan data tekstual yang bertujuan menemukan variabel tersembunyi. Dalam Penelitian ini akan fokus pada penggunaan pemodelan topik untuk mencari kemiripan antar pertanyaan dan jawaban. Hasil ekstraksi fitur ini akan dijadikan inputan untuk classfier untuk membuat model yang akan digunakan oleh data uji. Pada pengerjaan tugas akhir ini mengunakan Support Vector Machine (SVM) dan logistic regression untuk mendapatkan score klasifikasi dimana score ini yang menentukan peringkat sebuah jawaban untuk setiap pertanyaan. Perbedaan penelitian ini dengan penelitian milik JAIST adalah proses pemberian peringkat pada jawaban. Pada sistem milik JAIST hanya sampai proses klasifikasi saja belum ada sistem pemeringkatan jawaban.

Berdasarkan hasil evaluasi dari penelitian yang dilakukan penulis didapatkan bahwa sistem pemeringkatan yang dibuat memiliki nilai mean avarage precision sebesar 71.9%. Hasil ini didapatkan mengunakan logistic regression sebagai classifiernya. Jika dibandingkan dengan hasil SemEval 2016, hasil yang didapatkan dalam penelitian ini berada di rangking 8 dari 13 sistem yang dibuat peserta SemEval 2015 task 3 .

Kata Kunci: community question answering,tekstual fitur, pemodelan topik,ekstraksi fitur,Peringkat