

1. Pendahuluan

1.1. Latar Belakang

Di zaman modern ini manusia sudah semakin bergantung kepada internet, serta diikuti dengan kebutuhan akan informasi juga meningkat. Maka sistem *Community Question Answering* (CQA) akan membantu pengguna untuk menemukan informasi yang dibutuhkan, contohnya *Qatar Living Forum* yang merupakan sumber dari *dataset* yang disediakan SemEval-2017 Task 3. Selain itu CQA merupakan tempat umum yang bisa diakses oleh semua orang. Semua orang dapat mengajukan pertanyaan dan memberikan jawaban sesuai keinginan mereka sendiri sehingga untuk mendapatkan jawaban yang sesuai seperti yang diharapkan sangat sulit karena banyaknya jawaban di setiap pertanyaan dan tidak semua jawaban sesuai dengan pertanyaan yang ditanyakan. Terutama jika jawaban yang paling tepat ada di pertanyaan lain yang serupa dengan jumlah yang banyak. Maka dari itu akan memakan banyak waktu untuk mendapatkan jawaban yang tepat dan hal ini sangat tidak menguntungkan dan membuang-buang waktu.

Dalam sistem *Community Question Answering* maka banyak juga celah dari sistem yang dapat dikembangkan lagi. Terutama pada bagian pemilihan jawaban yang sangat susah untuk dilakukan secara *manual*, sehingga dibutuhkan otomasi yang dapat membantu pengguna memilah jawaban yang berkualitas dengan yang tidak berkualitas. Bantuan yang dapat diberikan berupa pemberian peringkat jawaban.

Pada tugas akhir ini akan mengembangkan sistem yang akan membantu pengguna CQA untuk mendapatkan jawaban yang relevan dari banyak jawaban dari setiap pertanyaan terkait, karena dengan banyaknya jawaban akan menambah waktu yang digunakan pengguna CQA dalam mencari jawaban yang tepat. Tetapi dengan adanya dua metode yang serupa yaitu *Cosine Similarity* dan *Soft-Cosine Semantic Similarity* (SCSS) akan dibandingkan untuk menentukan metode yang lebih membantu pengguna dalam mencari jawaban, karena SCSS merupakan metode yang belum lama muncul dan lebih digunakan untuk menangani kasus *similarity* antar pertanyaan [1] berbeda dengan *Cosine Similarity* yang sudah banyak digunakan untuk menangani berbagai kasus *similarity*. Data yang akan digunakan merupakan *dataset* yang disiapkan SemEval-2017 Task 3 yang berasal dari *Qatar Living Forum* dan berbahasa Inggris.

1.2. Topik dan Batasan

Berdasarkan deskripsi latar belakang yang sudah dipaparkan maka dapat dirumuskan beberapa masalah di tugas akhir ini seperti cara mengidentifikasi komentar yang relevan, membandingkan fitur yang mampu menghubungkan keterkaitan antar pertanyaan dengan komentar relevan, dan memberikan peringkat pada komentar *Good*. Dari semua permasalahan yang ada dibuat batasan dalam melakukan penelitian berupa data yang disiapkan oleh SemEval -2017 Task 3 yang berisi *dataset* dari *Qatar Living Forum*, lalu fitur yang digunakan hanya *Cosine Similarity* dan *Soft-Cosine Semantic Similarity* dan yang terakhir proses klasifikasi dalam pemberian peringkat dengan mengambil sepuluh kelas *good* teratas untuk menentukan kualitas.

1.3. Tujuan

Tujuan dari tugas akhir ini untuk menjawab permasalahan yang ada yaitu dapat mengidentifikasi komentar yang relevan dengan menggunakan fitur yang ditentukan dilanjutkan dengan mendapatkan fitur yang berdampak besar terhadap performa dan diakhiri dengan memberikan peringkat komentar sesuai hasil klasifikasi.

1.4. Organisasi Tulisan

Jurnal TA ini terdiri dari beberapa bagian setelah pendahuluan yaitu Studi Terkait, Sistem yang Dibangun, Evaluasi, dan Kesimpulan. Bagian studi terkait menjelaskan mengenai teori dasar yang digunakan dalam penanganan *question answering* seperti jenis-jenis kasus pada *community question answering*, data yang digunakan, penjelasan *praproses*, ekstraksi fitur, klasifikasi, pemberian peringkat, evaluasi performansi sistem dan *library* yang digunakan dalam penelitian. Bagian sistem yang dibangun menjelaskan mengenai rancangan sistem yang dibangun berdasarkan mekanisme dan batasan yang digunakan. Bagian evaluasi menjelaskan mengenai data yang digunakan, pengujian dan hasil pengujian. Analisis menggunakan hasil dari ekstraksi fitur dan klasifikasi. Bagian kesimpulan berisi kesimpulan terhadap pengaruh ekstraksi fitur dan klasifikasi terhadap nilai MAP yang dihasilkan. Saran sebagai masukan untuk penelitian lebih lanjut.