

# BAB 1

## PENDAHULUAN

Pada Bab 1 ini berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian dan sistematika penulisan. Berikut ini penjelasan dari masing-masing subbab tersebut.

### 1.1 Latar Belakang

Saat ini perkembangan teknologi sangatlah pesat terutama teknologi internet yang dapat membantu manusia mendapatkan informasi dengan cepat. Menurut data dari Miniwatts Marketing Group<sup>1</sup> pengguna internet dunia mencapai 3.270.490.584 yaitu sebesar 45% dari populasi manusia dunia. Internet pun mempengaruhi perilaku manusia baik dalam mendapatkan informasi atau memberi informasi. Pengguna Internet dapat menggunakan mesin pencari seperti google untuk mencari informasi tetapi salah satu kelemahan mesin pencari adalah saat penggunaanya menginputkan pertanyaan maka mesin pencari akan mencari dokumen yang relevan dengan pertanyaan tersebut tetapi tidak dapat memberikan jawaban secara langsung. Saat ini banyak komunitas tanya-jawab (*Community Question Answering*) seperti Qatar Living Forum yang dapat membantu penggunaanya mendapatkan informasi yang diinginkan. Setiap orang dapat bertanya maupun menjawab pada komunitas ini.

Pengguna dapat memuat pertanyaan mereka pada website CQA dan pengguna yang lain bebas memuat jawaban mereka terhadap pertanyaan tersebut. Sehingga satu pertanyaan dapat memiliki list jawaban yang sangat banyak dan beragam. Jawaban tersebut dapat memuat informasi yang memang relevan atau tidak relevan sama sekali dengan pertanyaan sehingga dibutuhkan tenaga dan usaha untuk memilih jawaban yang paling sesuai dengan pertanyaannya. Hal tersebut membutuhkan waktu yang lama jika dilakukan secara manual, maka dibutuhkan sistem yang secara otomatis memberikan peringkat sesuai kesesuaiannya jawaban dengan pertanyaan pada list jawaban tersebut.

SemEval 2015 task 3 subtask A1<sup>2</sup> telah menyelesaikan masalah tentang pemilihan jawaban yang paling sesuai dengan pertanyaan. Pada task ini dilakukan klasifikasi jawaban menjadi menjadi 3 kelas yaitu Good, Potentially Useful dan Bad. Contohnya dapat dilihat pada lampiran 1. Pada penelitian ini belum dapat menyelesaikan masalah pemeringkatan jawaban hanya memberi label pada jawaban saja.

Berdasarkan hal tersebut, penulis akan membangun sistem yang secara otomatis memberikan peringkat pada jawaban dari sebuah pertanyaan. Pemodelan topik merupakan perkembangan dari analisis teks yang bermanfaat dalam pemodelan data tekstual dengan tujuan menemukan topik yang tersembunyi di dalamnya. Jawaban yang bagus pasti memiliki topik yang sama atau mirip dengan pertanyaan. Tekstual fitur merupakan fitur yang didapatkan dengan melihat isi dari data teks itu sendiri. Tekstual fitur dapat menentukan kualitas jawaban karena informasi penting dari sebuah jawaban lebih

---

<sup>1</sup><http://www.statista.com/statistics/273018/number-of-internet-users-worldwide>

<sup>2</sup><http://alt.qcri.org/semeval2015/task3>

banyak terdapat pada isi teksnya. Sebagai contoh apakah sebuah jawaban mengandung tanda tanya atau simbol tertawa jika ya, maka jawaban tersebut dapat diklasifikasikan sebagai jawaban yang buruk. Pemodelan topik dan tekstual fitur akan digunakan sebagai parameter untuk proses klasifikasi. Klasifikasi dan pemeringkatanya akan menggunakan metode support vektor mesin (SVM) dan logistic regression. SVM dipilih karena merujuk pada hasil penelitian sebelumnya yaitu SemEval 2015 task 3. Penggunaan SVM-regression yang menghasilkan hasil terbaik. Sedangkan logistic regression digunakan karena melihat dari sejarahnya bahwa logistic regression merupakan awal mula terlahirnya SVM. Sistem ini diharapkan dapat memberikan peringkat pada jawaban sesuai dengan kualitas jawaban tersebut terhadap pertanyaannya. Sehingga jawaban dengan kualitas yang baik dapat di atas jawaban yang kualitas lebih buruh, serta dapat melihat fitur-fitur apa saja yang baik dalam memberikan peringkat pada jawaban.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang diatas, maka rumusan masalah yang akan dibahas adalah :

1. Bagaimana mengimplementasikan tekstual fitur dan pemodelan topik dalam memberikan peringkat pada list jawaban ?
2. Bagaimana pengaruh tekstual fitur dan pemodelan topik terhadap proses klasifikasi dan pemeringkatan jawaban.?
3. Bagaimana Pengaruh SVM dan logistic regression yang digunakan pada proses klasifikasi dan pemeringkatan jawaban?

## **1.3 Tujuan**

Pada tugas akhir ini mempunyai tujuan dan kegunaan sebagai berikut:

1. Merancang dan melakukan pemeringkatan jawaban dengan menggunakan tekstual fitur dan pemodelan topik,
2. Menganalisa pengaruh masing-masing fitur dan performasi sistem,
3. menganalisi pengaruh SVM dan logistic regression untuk klasifikasi dan pemeringkatan jawaban .

## **1.4 Batasan Masalah**

Pada Tugas Akhir ini akan dibuat beberapa batasan masalah untuk menghindari meluasnya materi pembahasan. Adapun batasan masalahnya mencakup hal-hal berikut:

1. Sistem yang dibangun hanya untuk dataset bahasa inggris yang diambil dari Qatar Living forum,
2. Sistem hanya terdiri dari tiga kelas Good ,PotentiallyUseful dan Bad,
3. Dibangun menggunakan bahasa pemrograman java.

## 1.5 Metodologi Penelitian

Adapun beberapa tahapan metode penelitian yang digunakan pada tugas akhir ini adalah sebagai berikut:

1. Studi literatur

Pada tahap ini akan dilakukan pengumpulan data dan informasi yang berkaitan dengan masalah-masalah yang terkait pada tugas akhir ini. Diantaranya adalah memahami konsep mengenai proses ekstraksi fitur dalam NLP, memahami SVM dan logistic regression untuk menentukan peringkat sebuah jawaban dan mempelajari *question answer* sistem. Hal-hal tersebut didapatkan dari buku, jurnal ilmiah, dan internet.

2. Analisis

Pada tahap ini dilakukan analisis terhadap data-data yang telah dikumpulkan melalui tahap studi literature, dimana data-data tersebut berupa kumpulan fitur yang dapat digunakan untuk menyelesaikan permasalahan pada tugas akhir ini yaitu masalah pemeringkatan jawaban. Dari analisis yang dilakukan akan menghasilkan fitur-fitur yang akan digunakan dalam pengerjaan tugas akhir. Pemilihan fitur ini melihat fitur-fitur yang paling berpengaruh pada penelitian sebelumnya.

3. Perancangan Sistem

Pada tahap ini dilakukan perancangan sistem pemeringkatan jawaban pada Community question answering(CQA). Secara umum sistem ini membantu menentukan peringkat jawaban dari sebuah pertanyaan. Sistem ini dirancang seperti mesin pencari, dimana user memilih pertanyaan lalu sistem akan menghasilkan jawaban dengan urutan jawaban yang telah diberi peringkat.

4. Implementasi Sistem

Pada tahap ini dilakukan pembuatan antarmuka dan coding. Implementasi aplikasi ini menggunakan bahasa pemrograman Java.

5. Pengujian sistem

Pengujian sistem dilakukan dengan melihat hasil evaluasi sistem yaitu berupa nilai MAP dan F1-measure serta memastikan kelayakan aplikasi yang dibuat.

6. Analisa Pengujian Sistem

Pada tahap ini dilakukan analisa data yang didapatkan dari hasil pengujian yang dilakukan sehingga dapat diketahui tingkat akurasi sistem dan kelayakan sistem.

7. Penulisan Laporan

Setelah menyelesaikan seluruh tahap pengembangan penelitian diatas, tahap selanjutnya adalah pembuatan laporan akhir dan dokumentasi yang sesuai dengan sistem yang dibuat. Laporan akhir ini dibuat dalam bentuk buku. Buku ini berisi seluruh dokumentasi sistem yang dibangun dari tahap awal pembuatan hingga tahap akhir pengujian secara lengkap dan terperinci.

## 1.6 Sistematika Penulisan

Secara garis besar penulisan laporan akhir penelitian ini terdiri dari 5 (lima) bab dengan beberapa sub bab. Agar mendapat arah dan gambaran yang jelas mengenai hal yang tertulis, berikut ini sistematika penulisannya secara lengkap:

### 1. BAB 1 PENDAHULUAN

Pada bab ini membahas latar belakang masalah, pembatasan dan perumusan masalah, tujuan penelitian, metodologi penelitian, dan sistematika penulisan laporan penelitian.

### 2. BAB 2 TINJAUAN PUSTAKA

Pada bab ini diuraikan tentang konsep *question answering* konsep analisis tekstual fitur, konsep pemodelan topik dan konsep Support Vector Machine (SVM) dan logistic regression. Serta penulis menjelaskan tentang konsep latent dirichlet allocation(LDA) dan yang terakhir konsep penilaian akurasi dalam proses klasifikasi dan sistem pemeringkatan.

### 3. BAB 3 PERANCANGAN SISTEM

Pada bab ini dijelaskan tentang gambaran umum penelitian yang akan dilakukan terdiri atas proses Preprocessing, ekstraksi fitur, klasifikasi jawaban dengan menggunakan metode Support Vector Machine (SVM), pemberian peringkat pada jawaban dan penentuan hasil pemeringkatan jawaban.

### 4. BAB 4 EVALUASI

Pada bab ini dijelaskan tentang pengujian dan analisis yang didapatkan dengan melakukan perbandingan terhadap beberapa fitur dan classifier.

### 5. BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan-kesimpulan yang didapat dari hasil penelitian dan berisi saran-saran yang sesuai dengan permasalahan yang diteliti.