

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi pada saat ini berkembang begitu pesat. Pada saat ini informasi dapat diperoleh dimana saja salah satunya melalui *website*. Salah satu fungsi dari *website* adalah sebagai media informasi [1]. Banyak *website* yang menyediakan informasi sesuai yang diharapkan oleh pembacanya termasuk juga mengenai informasi perguruan tinggi.

Perguruan tinggi memiliki informasi yang dapat disuguhkan kepada mahasiswa. Saat ini banyak Perguruan tinggi sudah mulai mendigitalisasi sistem akademiknya. Dari menyediakan informasi yang berkaitan dengan kampus, mahasiswa, dan dosen sampai teknis dari pembelajaran atau yang biasa disebut layanan akademik.

Pada saat ini informasi pada layanan akademik sudah cukup lengkap untuk disediakan. Namun terkadang orang khususnya mahasiswa tidak menemukan informasi yang diinginkannya dikarenakan informasi tersebut tidak ada atau kesalahan dari mahasiswa itu sendiri seperti salah mencantumkan kata atau tidak membaca informasi secara lengkap. Sebagaimana yang Allah Swt. perintahkan dalam Al-Qur'an pada surat An-Nahl ayat 43 yang berbunyi:

وَمَا أَرْسَلْنَا مِنْ قَبْلِكَ إِلَّا رِجَالًا نُوحِي إِلَيْهِمْ فَاسْأَلُوا أَهْلَ الذِّكْرِ إِنْ كُنْتُمْ لَا تَعْلَمُونَ ﴿٤٣﴾

“Dan Kami tidak mengutus sebelum kamu, kecuali orang-orang lelaki yang Kami beri wahyu kepada mereka, maka bertanyalah kepada orang yang mempunyai pengetahuan jika kamu tidak mengetahui” (QS. An-Nahl : 43). Pada ayat tersebut Allah SWT. memerintahkan untuk bertanya

kepada orang – orang yang kompeten atau berilmu jika tidak mengetahui [2].

Bertanya melalui sosial media menjadi salah satu cara mahasiswa untuk mencari informasi yang ingin didapat dengan harapan pertanyaan tersebut akan dijawab. Pertanyaan yang diajukan dapat berupa komentar yang diposting pada postingan akun resmi maupun berupa pesan langsung [3]. *Instagram* merupakan salah satu platform yang dijadikan tempat mahasiswa untuk mengajukan pertanyaan khususnya pada postingan akun *official*. Jika semua pertanyaan tersebut dijawab satu per satu maka akan membutuhkan waktu yang sangat lama. Hal ini menyebabkan kurang optimalnya sistem pada layanan akademik sehingga dibutuhkan otomasi dalam penyajian informasi pada layanan akademik tersebut.

Otomasi dalam penyajian informasi pada layanan akademik diharapkan dapat menyajikan informasi secara otomatis sesuai keinginan atau pertanyaan mahasiswa tanpa pihak kampus menjawabnya satu per satu. Hal ini dapat memberikan layanan yang lebih baik bagi mahasiswa dan meningkatkan efisiensi dalam proses administrasi akademik di perguruan tinggi. Otomasi ini dapat dilakukan dengan cara mengumpulkan pertanyaan mahasiswa berdasarkan frekuensinya dan mengekstraksi informasi dari layanan akademik yang relevan sebagai jawabannya.

Ekstraksi informasi adalah proses mengidentifikasi informasi yang relevan dari sejumlah besar dokumen atau data yang tersedia, dengan tujuan menemukan informasi yang diinginkan oleh pengguna [4]. Dalam otomasi layanan akademik, ekstraksi informasi dapat digunakan untuk mengambil informasi penting dari dokumen-dokumen akademik seperti jadwal kuliah, nilai mahasiswa, dan lain-lain, sehingga dapat mempermudah proses pengelolaan akademik. Namun tanpa adanya pertanyaan yang akan dijawab, ekstraksi informasi akan menjadi kurang efisien karena tidak ada informasi yang akan dicapai. Proses menemukan atau mengumpulkan pertanyaan dapat dilakukan salah satunya dengan cara mengambil dari media sosial.

Adapun cara untuk mengambil data dari internet yaitu dengan *Web Scraping*. *Web Scraping* adalah sebuah cara pengambilan suatu data atau informasi tertentu dengan jumlah besar untuk nantinya digunakan dalam berbagai keperluan seperti riset, analisis dan lainnya [5]. Namun ketika pertanyaan sudah terkumpul terkadang pertanyaan tersebut menggunakan kata tidak baku atau terdapat pertanyaan yang sama atau berulang. Sehingga harus dilakukan penelitian yang menghasilkan suatu atau beberapa pertanyaan unik dan bersifat baku.

Metode *N-Gram* digunakan untuk mengambil potongan kata sejumlah n dari sebuah rangkaian kata (kalimat, paragraf, bacaan) yang secara kontinuitas dibaca dari teks sumber hingga akhir dari dokumen [6] sedangkan *Term Frequency* merupakan kuantitas *term* yang sering muncul dalam suatu dokumen [7]. Metode *N-Gram* dapat digunakan untuk mencari bahan yang akan digunakan dalam pembangkit pertanyaan dan *Term Frequency* dapat digunakan untuk menghitung banyaknya data yang mengandung *N-Gram*. Data yang mengandung *N-Gram* dan memiliki frekuensi terbanyak nantinya akan dicari kemiripan dengan hasil pembangkit pertanyaan untuk menentukan apakah hasil *generate* sesuai atau tidak dengan data yang ada.

Pendekatan *Cosine Similarity* dapat menjadi solusi dalam menentukan apakah hasil *generate* sesuai atau tidak dengan data yang ada. *Cosine Similarity* adalah metode untuk menentukan tingkat kesamaan, yang akan menghitung sudut yang merepresentasikan vektor *query* dan vektor dokumen [8]. Dengan menggunakan metode *Cosine Similarity*, sistem dapat mengetahui tingkat kesesuaian antara pertanyaan yang dihasilkan dari pembangkit pertanyaan dengan data yang ada.

Berdasarkan uraian latar belakang diatas, untuk mengetahui apakah pendekatan *Cosine Similarity* dapat menentukan tingkat kesamaan atau kesesuaian pertanyaan dengan baik dan relevan dengan data yang ada maka perlu adanya penelitian. Dengan permasalahan tersebut maka dirumuskan sebuah penelitian yang berjudul “**Kesesuaian Pembangkit**

Pertanyaan Untuk Layanan Akademik Dengan *Cosine Similarity*”, penelitian ini diharapkan menghasilkan model yang dapat membuat pertanyaan unik serta baku dan relevan dengan data yang ada.

1.2 Rumusan Masalah

Rumusan masalah didasarkan pada latar belakang adalah sebagai berikut:

1. Bagaimana membangkitkan pertanyaan berdasarkan data yang ada?
2. Bagaimana kinerja *Cosine Similarity* dalam kesesuaian pembangkit pertanyaan untuk layanan akademik?
3. Bagaimana menampilkan hasil pembangkit pertanyaan yang sesuai ke dalam sebuah aplikasi web?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Melakukan pembangkitan pertanyaan berdasarkan data yang ada.
2. Mengetahui kinerja *Cosine Similarity* dalam kesesuaian pembangkit pertanyaan untuk layanan akademik.
3. Menampilkan pertanyaan hasil pembangkitan dalam sebuah aplikasi web.

1.4 Manfaat Penelitian

Manfaat yang didapat dari penelitian ini adalah sebagai berikut:

1. Memudahkan mahasiswa mencari pertanyaan yang sering diajukan.
2. Memudahkan peneliti selanjutnya untuk membuat model *Information Retrieval* berdasarkan pertanyaan yang sudah di *generate* dari penelitian ini.
3. Dapat menjadi bahan referensi bagi peneliti berikutnya mengenai pembangkit pertanyaan baik menggunakan konsep *Cosine Similarity* maupun yang lainnya.

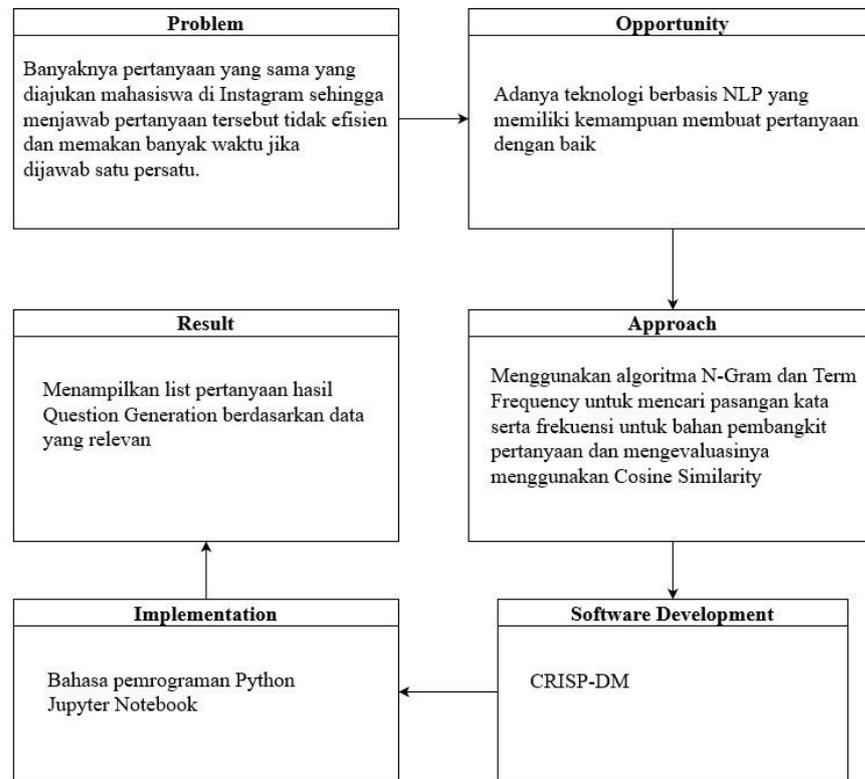
1.5 Batasan Masalah

Batasan masalah dari penelitian ini bertujuan agar penelitian ini menjadi lebih terarah dan sesuai dengan tujuan yang diinginkan. Berikut merupakan beberapa batasan masalah dari penelitian mengenai kesesuaian pembangkit pertanyaan untuk layanan akademik dengan *cosine similarity* adalah:

1. Sumber data yang digunakan pada penelitian ini yaitu dari media sosial *Instagram*.
2. Data yang diambil berupa komentar pada postingan yang di posting pada bulan September 2022 sampai Mei 2023 oleh akun *uinsgd.official* dengan total data 5797 komentar.
3. Sistem hanya dapat menerima masukan berupa teks berupa komentar.
4. Metode atau algoritma yang digunakan pada penelitian ini adalah menggunakan *Cosine Similarity* dengan bantuan *N-Gram* dan *Term Frequency* dalam mencari kata terbanyak.
5. Pembangkit pertanyaan dihasilkan dengan aturan *WH-Questions*.
6. Hasil akhir sistem menampilkan *list* hasil pembangkit pertanyaan yang relevan dengan data.
7. Pertanyaan yang dapat ditampilkan berdasarkan data yang diambil maksimal satu hari terakhir.

1.6 Kerangka Pemikiran

Kerangka pemikiran pada penelitian ini ditunjukkan pada gambar berikut:



Gambar 1. 1 Kerangka Pemikiran

1.7 Sistematika Penulisan

Sistematika penulisan pada penelitian ini dibagi ke dalam 5 bab. Pada setiap bab dijelaskan sesuai tujuan dan pengembangan dari sistem tersendiri. Sistematika penulisan pada penelitian ini dapat dilihat sebagai berikut:

BAB I : Pendahuluan

Bab I berisikan tentang latar belakang penelitian, rumusan masalah, tujuan, batasan, metode pengembangan sistem hingga kerangka pemikiran. Berikut dengan sistematika penulisan yang disajikan.

BAB II : Kajian Literatur

Bab II menjelaskan tentang pembahasan penelitian terdahulu serta konsep – konsep dan teori pendukung pada penelitian yang akan dilakukan.

BAB III : Metodologi Penelitian

Bab III berisikan tentang metode yang digunakan dalam penyusunan tugas akhir. Metodologi penelitian disajikan berdasarkan analisis kebutuhan menggunakan metode *CRISP-DM*. Dalam metode tersebut terdapat beberapa tahapan yaitu Pemahaman Bisnis, Pemahaman Data, Persiapan Data, Pemodelan, Evaluasi, dan *Deployment*.

BAB IV : Hasil dan Pembahasan

Bab ini membahas mengenai hasil dari implementasi sistem itu sendiri, memaparkan hasil penelitian serta pembahasan dari apa yang dilakukan.

BAB V : Simpulan dan Saran

Bab ini berisi kesimpulan dari penelitian yang dilakukan serta saran yang direkomendasikan untuk peningkatan atau perbaikan dari penelitian ini.

