

BAB I PENDAHULUAN

1.1 Latar Belakang

Kebutuhan informasi menjadi sebuah kebutuhan pokok seiring dengan berkembangnya teknologi informasi. Semakin bertambahnya informasi mengakibatkan dibutuhkan mesin pencari yang dapat mengakses dan menyediakan berbagai informasi. Mesin pencari tersebut mampu menyediakan informasi yang dibutuhkan oleh *user* dan menghasilkan informasi yang akurat dan relevan.

Penggunaan mesin pencari sudah banyak dipakai dalam dunia online namun hasil pencarian seringkali berbeda tergantung dari metode yang digunakan. Mesin pencari tersebut melakukan pencarian berdasarkan *query* yang dimasukkan oleh *user*, kemudian mencocokkan *query* dengan data yang tersedia[1]. Informasi yang dihasilkan oleh mesin pencarian bisa saja tidak relevan dengan yang diinginkan oleh *user*. Hal tersebut bisa terjadi karena *user* tidak mengerti *query* yang harus dimasukkan. Sehingga *user* harus memasukkan *query* kembali yang sesuai dengan yang diinginkan. Sehubungan dengan keterbatasan sumber daya komputasi dan waktu, maka dibutuhkan suatu cara untuk meningkatkan *query* sehingga dalam proses pencarian menghasilkan informasi yang lebih relevan tanpa *user* harus memikirkan kembali *query* yang baru.

Relevance feedback atau juga dikenal sebagai RF merupakan proses optimasi *query*, yang meningkatkan jumlah dokumen yang relevan dari hasil uji coba pencarian[2]. Proses *relevance feedback* akan mengembalikan dokumen hasil pencarian yang mempunyai tingkat kesamaan dengan dokumen yang dicari. Dengan *relevance feedback*, *user* dapat memberikan *feedback* untuk mengganti *query* dengan yang lebih spesifik sehingga akan didapatkan hasil yang lebih relevan.

Salah satu metode yang digunakan pada *Relevance feedback* yaitu metode *Rocchio*. Metode *Rocchio relevance feedback* merupakan strategi reformulasi *query* paling populer karena sering digunakan untuk membantu *user* dalam menyelesaikan suatu *information retrieval systems*. Seperti yang digunakan pada penelitian similaritas pencarian dokumen tugas akhir menggunakan metode *Rocchio*[3].

Dalam penelitian studi kelayakan pencarian buku, dirancang aplikasi mobile pencarian buku yang juga dipadukan dengan metode *Rocchio*. Metode tersebut memanfaatkan *request* dari *user* untuk membantu proses pencarian dokumen relevan. Kesimpulan didapatkan berdasarkan hasil perhitungan tingkat similaritas antara keyword dan term pada judul dokumen. Hasil yang diperoleh menunjukkan bahwa metode *Rocchio* yang digunakan memberikan hasil yang relevan dengan query yang telah dimasukkan oleh *user*[4].

Pada sistem pencarian informasi, selain menggunakan metode *rocchio*, juga dapat menggunakan algoritma genetika. Algoritma genetika merupakan algoritma yang menghasilkan solusi perkiraan dalam menyelesaikan masalah melalui seleksi genetik dan digunakan sebagai metode optimasi. Pada algoritma genetika, beberapa proses menggunakan sistem data yang diambil secara acak, namun dapat diatur tingkat pengacakan dan tingkat kontrol[2].

Algoritma genetika biasanya digunakan untuk menyelesaikan masalah penjadwalan mata pelajaran, algoritma genetika memiliki performa yang cukup baik dalam menemukan solusi yang optimal pada masalah penjadwalan. Pemilihan operator dalam algoritma genetika juga sangat menentukan keberhasilan dalam menemukan solusi yang optimal. Dalam algoritma genetika, terdapat perhitungan nilai fitness. Nilai fitness tersebut yang akan menyatakan baik atau tidaknya suatu solusi tersebut. Algoritma genetika memiliki tujuan untuk mendapatkan solusi berdasarkan nilai fitness yang paling tinggi. Dari hasil analisis ditemukan bahwa *performance* dari algoritma genetika cukup baik dalam menyelesaikan masalah penjadwalan mata pelajaran[5].

Algoritma genetika tidak hanya dapat digunakan untuk penjadwalan mata pelajaran, namun juga dapat digunakan untuk meningkatkan hasil dari *information retrieval*. Karena algoritma genetika telah menunjukkan hasil yang baik pada proses optimasi, sehingga dimasukkan sebagai bagian dari proses pencarian informasi[2]. Algoritma Genetika diterapkan dalam *information retrieval* untuk membentuk *formula query* baru melalui *relevance feedback* yang digunakan untuk pengambilan dokumen yang dianggap paling relevan.

Berdasarkan pada penelitian sebelumnya mengenai penggunaan algoritma genetika pada sistem temu kembali informasi yang dilakukan Salman Dziyaul

Azmi, menunjukkan penerapan *relevance feedback* dengan metode *roulletwheel selection* dan rekombinasi *crossover*. Matrik evaluasi yang digunakan adalah *Mean Average Precision* (MAP) dan *recall* rata-rata berdasarkan penilaian *user*. Pada penelitian ini, hasilnya menunjukkan peningkatan 15,2% dan 28,6% pada nilai MAP dengan menggunakan dua dataset dokumen bahasa Indonesia, yaitu tesis abstrak dan dataset berita[2].

Penelitian yang dilakukan oleh Cut Lilis Setiawati untuk menyelesaikan permasalahan *Travelling Salesman Problem*(TSP) dengan studi pengaruh hibridisasi *Roulette Wheel selection* dengan *Tournament Selection*. Tujuan dari penelitian ini adalah membandingkan metode *Roulette Wheel Selection* dengan *Tournament Selection* pada algoritma genetika. Hasil penelitian menunjukkan bahwa implementasi *Turnament Selection* memiliki nilai lebih optimal dibandingkan dengan *Roulette Wheel Selection*, perbandingan nya sebesar 44% lebih bagus dengan menggunakan *Tournament Selection*[6].

Dengan demikian, untuk meningkatkan sistem yang efektif dalam menghasilkan dokumen yang relevan dan akurat dengan memanfaatkan *feedback* dari *user*, dan juga untuk melihat sejauh mana pengaruh peningkatan kemampuan dari *relevance feedback* dengan perbandingan *roulletwheel selection* dan *tournament selection*. Dimana pada *roulletwheel selection* menggunakan nilai probabilitas kumulatif dan *range* sebelum dilakukan proses random, sedangkan pada *tournament selection* hanya melakukan proses random namun dengan random yang berulang kali, maka dibuatlah penelitian ini dengan tujuan dapat menghasilkan sistem yang lebih akurat dalam pencarian kemiripan dokumen.

1.2 Rumusan Masalah

1.2.1. Permasalahan

Dalam mesin pencarian, tidak semua informasi yang dihasilkan oleh sistem sesuai dengan yang diinginkan oleh *user*. Hal tersebut dikarenakan *user* memasukkan *query* yang kurang spesifik. Sehingga diperlukan *query* yang baru dari proses *Relevance Feedback* untuk memperoleh informasi yang relevan, yang diharapkan oleh *user*.

1.2.2. Usulan Solusi

Diusulkan penerapan algoritma genetika pada *Relevance feedback* untuk meningkatkan kemampuan pencarian dengan penggunaan *Rouletewheel selection* dan *Tournament Selection* pada proses algoritma genetika.

1.2.3. Pertanyaan Penelitian

Untuk mencapai tujuan pengujian yang diinginkan maka diperlukan pertanyaan penelitian. Pertanyaan tersebut sebagai berikut :

1. Berapa waktu dan akurasi yang diperlukan untuk memproses dokumen relevan jika menggunakan *rouletewheel selection* ?
2. Berapa waktu dan akurasi yang diperlukan untuk memproses dokumen relevan jika menggunakan *tournament selection* ?

1.3 Tujuan dan Manfaat

1.3.1. Tujuan

Penerapan algoritma genetika pada *relevance feedback* untuk mendapatkan dokumen yang relevan dan juga membandingkan metode *rouletewheel selection* dan *tournament selection*.

1.3.2. Manfaat

Dihasilkan sistem *relevance feedback* sehingga sistem dapat melakukan interaksi dengan *user* untuk mendapatkan dokumen yang lebih relevan

1.4 Batasan Masalah

Dalam pembuatan sistem ini, terdapat beberapa batasan masalah, antara lain :

1. Dokumen yang digunakan dalam penelitian sebanyak 1000 dokumen. Koleksi dokumen (*corpus*) yang digunakan adalah skripsi dari Portal Tugas Akhir (<https://pta.trunojoyo.ac.id/>)
2. Teks yang digunakan sebagai dokumen uji merupakan dokumen yang menggunakan bahasa Indonesia yang baku.
3. Sistem yang dibangun menggunakan bahasa pemrograman *Python*.

1.5 Sistematika Laporan

Sistematika dalam penulisan Proposal ini terdiri dari beberapa bab, yaitu :

BAB I PENDAHULUAN

Pada bab ini membahas tentang latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, serta sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini membahas tentang landasan teori dan penelitian sebelumnya.

BAB III METODE USULAN

Pada bab ini membahas tentang metode yang diusulkan (algoritma), arsitektur (*Input, Proses, Output*), Dataset, tahapan penelitian, serta rencana pengujian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini memaparkan dan menganalisis data-data yang didapatkan dari hasil pengujian.

BAB V KESIMPULAN

Pada bab ini menjelaskan mengenai kesimpulan akhir penelitian dan saran-saran yang direkomendasikan berdasarkan hasil pengujian untuk perbaikan proses selanjutnya.