

BAB I

PENDAHULUAN

1.1 Latar Belakang

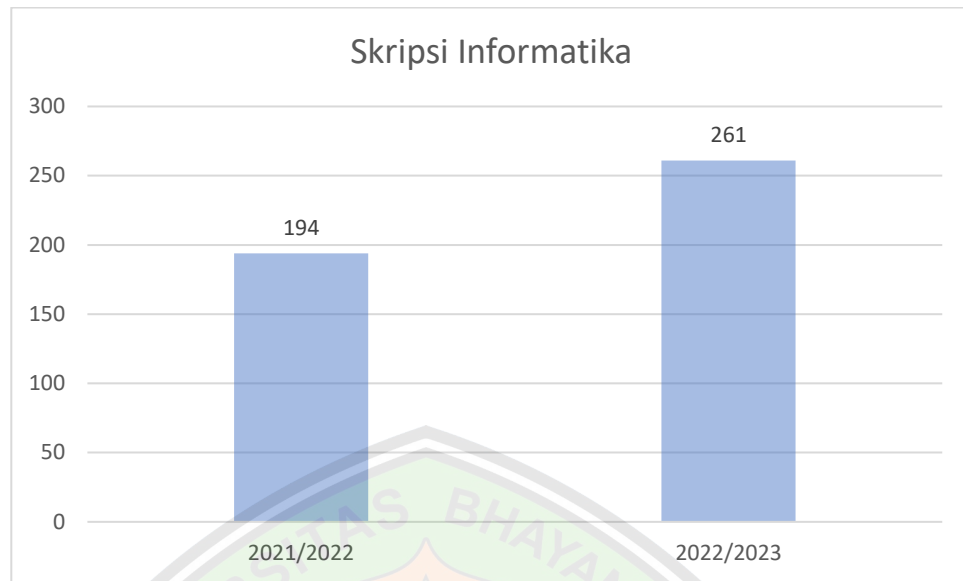
Universitas Bhayangkara Jakarta Raya (Ubhara Jaya) sebagai salah satu Perguruan Tinggi Swasta yang berada dibawah pembinaan Yayasan Brata Bhakti sebagai badan penyelenggaranya, berkewajiban mewujudkan tujuan pendidikan nasional dengan Visi dan Misi nya bagi mendukung keberhasilan tugas Kepoisian Negara Republik Indonesia khususnya dan pengembangan kuitas hidup bermasyarakat berbangsa dan bernegara pada umumnya.

Similarity Detection atau deteksi kemiripan merupakan salah satu implementasi dari NLP. Tujuan dari *similariry detection* adalah untuk dapat mengukur tingkat kemiripan dua buah objek. Salah satu objek yang bisa digunakan pada *Similarity Detection* adalah dokumen teks. Dokumen teks merupakan sebuah tulisan yang tercetak dan bertujuan untuk menerangkan atau memberikan sebuah informasi tertentu [1]. Beberapa metode *similarity detection* yang dapat menghitung persentase kemiripan dokumen adalah metode *Hamming*, *Demarau Levenshtein*, *Levenshtein Distence*, dan *Jaro-Winkler Distance* [2]. Terdapat beberapa penelitian *similarity detection* yang sudah dilakukan. Pada penelitian tahun 2022 metode *Jaro-Winkler Disttance* guna mengukur tingkat kemiripan pada dokumen Perundang-Undangan. Hasil dari uji coba tersebut yaitu perhitungan Jaro

Winkler dapat membandingkan kalimat yang mengalami perubahan struktur kata, namun pada perhitungan Levenshtein Distance hanya dapat bekerja lebih baik bila ada dua kalimat yang dianalisis memiliki struktur kata/kalimat yang serupa [3],[4].

Pada penelitian kali ini metode yang digunakan untuk nilai persentase *similarity* adalah metode *Jaro-Winkler Distance*. Dalam metode *Distance*, semakin tinggi nilai dari *Jaro-Winkler Distance* untuk dua kata pada dokumen maka semakin tinggi juga nilai dari *Similarity* kedua dokumen tersebut. Penelitian pada tahun 2022 yang membandingkan antara metode *Jaro-Winkler Distance* dengan *Levenshtein Distance* menyatakan bahwa perhitungan *Jaro-winkler Distance* memiliki tingkat *Similarity* lebih baik daripada menggunakan *Levenshtein Distance* [3].

Pada Universitas Bhayangkara Jakarta Raya memiliki dua perpustakaan yang terletak di kampus Jakarta dan kampus Bekasi. Keberadaan perpustakaan Universitas Bhayangkara Jakarta Raya (Ubhara Jaya) adalah sejalan dengan keberadaan Ubhara itu sendiri. Dokumen yang digunakan pada penelitian ini adalah abstrak skripsi program studi fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya sebanyak 455 abstrak skripsi. Dimana pada Universitas Bhayangkara Jakarta Raya (Ubahara Jaya), skripsi merupakan salah satu persyaratan dasar untuk gelar sarjana. Banyak mahasiswa menganggap skripsi sangat sulit. Sehingga banyak mahasiswa yang sengaja menyalin karya orang lain karena tidak paham dengan materi kuliah yang diajarkan, sengaja menulis karya orang lain karena tidak tahu bagaimana cara mengutip dan mencantumkan sumber informasi dengan baik dan benar. Berikut merupakan diagram batang skripsi informatika fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya 2 tahun terakhir:



Gambar 1. 1 Skripsi Informatika

Sumber: Hasil Penelitian (2024)

Gambar 1.1 merupakan diagram dataset yang digunakan pada penelitian ini yaitu menggunakan abstrak skripsi informatika fakultas ilmu komputer Universitas Bhayangkara Jakarta Raya rentang waktu 2 tahun terakhir. Berikut merupakan contoh dataset abstrak skripsi yang digunakan pada penelitian ini:

Tabel 1.1 contoh dataset abstrak skripsi

Dataset Abstrak Skripsi
<p><i>Indeks Pembangunan Manusia (HDI) digunakan sebagai parameter statistik untuk mengevaluasi kemajuan dan kualitas hidup manusia di suatu negara. Rencana pembangunan manusia diukur melalui tiga aspek dasar: umur panjang dan sehat, pengetahuan, dan standar hidup yang layak. Penelitian tersebut berfokus pada penerapan algoritma k-means untuk mengidentifikasi pola dan mengelompokkan 34 provinsi di Indonesia berdasarkan indikator pembentuk IPM tahun 2022. Indikator-indikator tersebut meliputi Angka Harapan Hidup (UHH), Angka Harapan Hidup Sekolah (HLS), Rata-Rata Waktu Sekolah (RLS), dan Pengeluaran. Pengelompokan IPM ini bertujuan untuk mengidentifikasi variabel-variabel IPM yang sebaiknya menjadi prioritas dalam pembangunan. Hasil analisis menunjukkan terbentuknya dua kelompok (cluster) melalui K-Means Cluster Analysis. Cluster 1 mempunyai karakteristik provinsi dengan nilai tinggi hingga sangat tinggi pada UHH, HLS, RLS, dan output yang disesuaikan. Sedangkan Klaster 2 terdiri dari provinsi dengan nilai UHH, HLS, RLS, dan output yang disesuaikan sedang hingga tinggi .</i></p>

Berlandasan dari penjabaran latar belakang diatas maka penulis akan menerapkan Metode *Jaro-Winkler Distance* pada penelitian ini sesuai dengan judul **“METODE *JARO-WINKLER DISTANCE* UNTUK DETEKSI KEMIRIPAN ABSTRAK SKRIPSI PADA UNIVERSITAS BHAYANGKARA JAKARTA RAYA”**

1.2 Identifikasi Masalah

Berdasarkan latar belakang di atas, dapat diidentifikasi masalah sebagai berikut:

1. Masih banyaknya abstrak skripsi yang serupa.
2. Di perlukan sebuah sistem untuk mengetahui nilai *similarity* antar dokumen abstrak skripsi.

1.3 Rumusan masalah

Berdasarkan permasalahan yang ada penelitian ini menerapkan metode *Jaro-Winkler Distance* untuk mengetahui nilai *similarity* dokumen abstrak skripsi.

1.4 Batasan Masalah

1. Data yang digunakan adalah dokumen abstrak skripsi Bahasa Indonesia
2. Dokumen Fakultas Ilmu Komputer tanpa kata kunci yang diperoleh dari perpustakaan Universitas Bhayangkara Jakarta Raya dengan rentang waktu 2021-2023 dengan jumlah 455 abstrak skripsi.

1.5 Tujuan dan Manfaat Penelitian

1.5.1 Tujuan Penelitian

1. Mengetahui nilai *similarity* metode *Jaro-Winkler Distance* dalam mendeteksi kemiripan abstrak skripsi.
2. Membuat sistem untuk mendeteksi kemiripan abstrak skripsi.

1.5.2 Manfaat Penelitian

1. Mengetahui penggunaan Metode *Jaro-Winkler Distance* dapat menjadi pertimbangan untuk meningkatkan pengukuran kemiripan pada penelitian selanjutnya
2. Dapat bermanfaat bagi akademik atau masyarakat untuk membantu mendeteksi kemiripan dokumen abstrak skripsi

1.6 Sistematika Penulisan

Dalam sistematika penulisan ini bertujuan untuk membantu meringkas permasalahan penelitian dan mempermudah pembaca dalam memahami isi penelitian. Berikut pembahasan masing-masing bab:

BAB I PENDAHULUAN

Pada bab ini menjelaskan mengenai latar belakang penulisan laporan, identifikasi masalah, tujuan penelitian, manfaat penelitian, metode penelitian, sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini menjelaskan mengenai teori-teori yang berkaitan dengan topik pembahasan sebagai acuan dan

penunangan untuk melakukan penelitian dan perancangan sistem.

BAB III METODELOGI PENELITIAN

Bab ini akan menguraikan tentang objek penelitian, prosedur dan spesifikasi sistem berjalan, permasalahan, alternatif pemecah masalah, dan metode penelitian.

BAB IV PERANCANGAN SISTEM DAN IMPLEMENTASI

Bab ini menjelaskan tentang hasil dari pengkodean yang dilakukan, sehingga yang dibahas pada bab ini adalah bagaimana tampilan siste saat dijalankan. Selanjutnya di paparkan bagaimana aplikasi mobile ini di uji.

BAB V PENUTUP

Pada bab ini berisi tentang kesimpulan dan saran yang diharapkan dalam pengembangan sebuah