

BAB I

PENDAHULUAN

1.1. Latar Belakang

Data akademik merupakan data krusial yang dimiliki oleh pihak universitas. Salah satu dari data akademik yang dibutuhkan oleh universitas yaitu daftar nama Sekolah Menengah Atas atau Sederajat yang biasanya digunakan sebagai informasi identitas asal sekolah mahasiswa pada Sistem Akademik (SIKAD) ataupun Sistem Seleksi Penerimaan Mahasiswa Baru (SPMB).

Pada website resmi Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) menjelaskan salah satu poin mengenai prinsip seleksi mahasiswa perguruan tinggi negeri yaitu “Seleksi dilaksanakan berdasarkan prinsip memperhitungkan rekam jejak kinerja sekolah, antara lain: akreditasi sekolah, prestasi mahasiswa alumni sekolah bersangkutan, jumlah siswa yang diterima melalui jalur SNMPTN, SBMPTN dan Seleksi Mandiri tahun sebelumnya, serta prestasi lainnya yang ditentukan oleh masing-masing PTN” (SNMPTN, 2015). Prinsip tersebut kemudian dipertimbangkan untuk dijadikan sebagai salah satu poin seleksi pada sistem SPMB yang dilaksanakan oleh universitas. Untuk mengakomodir prinsip tersebut, pihak universitas harus memiliki sebuah database data SMA, MA, SMK dan MAK beserta nilai akreditasi dari masing-masing sekolah.

Untuk mendapatkan data sekolah tersebut, universitas dapat melakukan kerjasama dengan pihak lain yang mengelola data akreditasi sekolah. Salah satu pengelola data akreditasi sekolah bertaraf nasional yaitu Badan Akreditasi Nasional Sekolah/Madrasah (BAN-SM). Berdasarkan Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 59 Tahun 2012 Tentang Badan Akreditasi Nasional Sekolah/Madrasah Pasal 1 Ayat 2, dijelaskan bahwa BAN-SM adalah badan evaluasi mandiri yang menetapkan kelayakan program dan satuan pendidikan jenjang pendidikan dasar dan menengah jalur formal dengan mengacu pada standar nasional pendidikan. BAN-SM mempublikasikan semua data akreditasi sekolah mulai dari jenjang Taman Kanak-kanak (TK) hingga SMA/SMK

secara terbuka di situs resminya di bansm.or.id. Namun pihak BAN-SM tidak memberikan fasilitas koneksi langsung dengan database atau memberikan fasilitas *web service* untuk menggunakan data akreditasi tersebut. Cara yang paling mudah yang dapat dilakukan yaitu mengambil data-data yang diperlukan dari website bansm.or.id yaitu dengan menggunakan *crawling web* dan *web mining*.

Permasalahannya data yang terdapat di website bansm.or.id tidak memiliki *unique keywords* yang dapat digunakan sebagai acuan untuk merelasikan tabel sekolah milik universitas dengan data yang diambil dari website bansm.or.id, sehingga perlu mencocokkan nama sekolah yang ada di tabel sekolah milik universitas dengan data yang ada di website bansm.or.id. Masalah lainnya terdapat perbedaan susunan kata atau kesalahan penamaan sekolah yang ada di database milik universitas dengan data yang ada di website bansm.or.id, sehingga diperlukan algoritma *similarity* untuk melakukan pengecekan terhadap nama sekolah yang ada di tabel sekolah milik universitas dengan nama sekolah dari website bansm.or.id.

Berdasarkan penelitian Varol dan Bayrak tahun 2011 yang berjudul *Risk Assessment of Error-Prone Personal Information in Data Quality Tools* mengenai koreksi data nama pelanggan yang salah ejaan menggunakan Algoritma *Personal Name Recognition Strategy (PNRS)* menghasilkan tingkat koreksi terbaik sebesar 64,90%. Akan tetapi *PNRS* hanya mampu mengkoreksi nama-nama yang salah ejaan berdasarkan daftar nama-nama populer yang sering digunakan di Amerika Serikat. Untuk mengatasinya, digunakanlah algoritma *Jaro-Winkler Distance* yang menghasilkan tingkat koreksi terbaik kedua sebesar 62,1% - 63,6% (Varol & Bayrak, 2011). Hu, Zhang, & Wu mengusulkan algoritma *Hamming Distance based Approximate Similarity Text Search (HASTS)* untuk meningkatkan kualitas *query* pada data text yang besar. Hasil penelitiannya menunjukkan bahwa algoritma HASTS dapat melakukan pencarian *query* pada data teks yang besar dengan waktu dan *space* yang relatif rendah (Hu, Zhang, & Wu, 2015).

Karena sulit untuk menyediakan nama-nama populer yang benar sesuai dengan karakter penamaan pada nama sekolah maka akan sulit untuk memanfaatkan algoritma *PNRS*. Karena itu, penulis melakukan penelitian

mengenai pencocokan data sekolah berdasarkan nama sekolah menggunakan teknik *similarity* dengan algoritma *Jaro-Winkler Distance* dan *Hamming Distance*.

1.2. Rumusan Masalah

Berdasarkan penjabaran pada bagian latar belakang, masalah yang di selesaikan adalah bagaimana cara mencocokkan data kabupaten dan sekolah menggunakan algoritma *Hamming Distance* dan *Jaro-Winkler Distance*.

1.3. Batasan Masalah

Batasan masalah yang diterapkan pada penelitian ini sebagai berikut

1. Data yang digunakan hanya tabel sekolah pada database universitas dan data yang tersedia di *website* bansm.or.id.
2. Data akreditasi yang digunakan adalah data akreditasi SMA, MA, SMK dan MAK.
3. Tabel propinsi dari database universitas telah direlasikan dengan tabel propinsi yang dibuat berdasarkan pada daftar propinsi dari *website* bansm.or.id.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mendapatkan nilai akreditasi SMA/SMK dan mengukur akurasi dari hasil pencocokan sehingga diperoleh rekomendasi data akreditasi yang sesuai dengan data sekolah yang dimiliki universitas.

1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini yaitu dapat menyediakan tabel relasi antara tabel sekolah dari database universitas dengan tabel sekolah dari hasil *crawling website* bansm.or.id.

1.6. Sistematika Penulisan

Sistematika penulisan memuat tentang metode penulisan yang digunakan dalam pembuatan laporan penelitian, dengan uraian sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan laporan.

BAB II TINJAUAN PUSTAKA

Pada bab ini diuraikan secara umum teori-teori dan penelitian-penelitian yang berhubungan dengan materi yang nantinya digunakan untuk melakukan proses analisa dan implementasi aplikasi.

BAB III METODOLOGI

Pada bab ini diuraikan secara umum tentang gambaran objek penelitian, serta gambaran langkah-langkah yang dilakukan oleh penulis untuk melaksanakan dan menyelesaikan penelitian ini.

BAB IV PEMBAHASAN

Pada bab ini, dipaparkan hasil-hasil dari tahapan penelitian, dari tahap analisis, implementasi dan pengujian.

BAB V PENUTUP

Pada bab ini berisi kesimpulan dan saran. Kesimpulan berisi jawaban dari perumusan masalah. Sedangkan saran berisi pemecahan masalah yang diteliti atau tindak lanjut dari hasil penelitian.