

## BAB II

### TINJAUAN PUSTAKA

Pamutha et al., (2012) melakukan penelitian yang berfokus pada *preprocessing metode file log web* yang dapat digunakan untuk mengidentifikasi tugas sesi dari *web log*. Penelitian ini juga menghasilkan informasi statistik dari sesi pengguna, seperti: (1) jumlah IP yang unik; (2) total halaman unik; (3) Jumlah sesi; (4) panjang Sesi dan (5) frekuensi mengunjungi berdasarkan usia.

Meinanda et al., (2009) melakukan penelitian untuk memprediksi masa studi sarjana dengan *artificial neural network*. Dalam penelitian ini menggunakan *preprocessing data* agar pada tahap pembuatan model mampu menghasilkan model yang efektif. Beberapa hal yang dilakukan pada tahap data *preprocessing* adalah transformasi ke dalam bentuk yang lebih informatif dengan menggunakan *cross-tabulation*. Tahap data *preprocessing* selanjutnya adalah melakukan penghapusan terhadap *missing-value* dan *outlier* pada data.

Adriyani, et al., (2012) membuat sistem Pengecekan Ejaan Bahasa Indonesia yang dikembangkan dengan berbasis web menggunakan bahasa pemrograman PHP . Algoritma yang digunakan untuk memberikan saran perbaikan adalah Levenshtein Distance yang dapat menghitung keterkaitan antar string dan menghitung jumlah keterbedaan antar dua string. Disamping itu kesalahan penulisan juga dapat disebabkan kurangnya spasi antar kata sehingga kata tersebut tidak mengandung makna maka diterapkan pula metode empiris.

Junedy.S, (2014) menggunakan *levenshtein distance* untuk membuat sistem pendeteksian plagiarisme, yaitu untuk mendeteksi kemiripan dokumen teks. Algoritma *levenshtein distance* mampu mendeteksi kemiripan dokumen teks sehingga dapat digunakan untuk membantu menentukan plagiarisme. Untuk meningkatkan kerja algoritma *levenshtein distance* maka sebelumnya dilakukan preprocessing terhadap dokumen teks yang akan diproses. Dari ujicoba yang dilakukan, algoritma *levenshtein distance* mampu mendeteksi kemiripan dokumen teks, ini dibuktikan dengan nilai *similarity* yang tinggi untuk dokumen yang plagiat. Preprocessing terbukti dapat meningkatkan kinerja algoritma *levenshtein distance* meskipun menambah waktu untuk pemrosesan.

Berdasarkan hasil penelitian terdahulu diatas maka sistem pengolahan data jumlah calon mahasiswa baru yang penulis buat akan menggunakan metode *preprocessing data* dan didukung oleh algoritma *levenshtein distance*. Preprocessing data ini akan digunakan dalam pembersihan data yang kotor dan algoritma *levenshtein distance* akan membantu proses pembersihan data yaitu dalam pembersihan data nama sekolah. Dengan penggunaan metode ini diharapkan hasil yang dihasilkan akan membantu staf KKP dalam pengolahan data jumlah calon mahasiswa.