

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan dari penelitian “Pemanfaatan teknik *Semi-Supervised Learning* untuk klasifikasi dokumen medis”, diantaranya adalah:

- a. Klasifikasi dokumen medis menggunakan teknik *Semi-Supervised Learning* memanfaatkan dua jenis data yaitu data berlabel dan data tidak berlabel. Berdasarkan penelitian yang dilakukan, klasifikasi memanfaatkan data berlabel menggunakan metode *Multinomial Naïve Bayes*, dan memanfaatkan metode *Pseudo Labeling* pada data tidak berlabel. Pemanfaatan dua metode tersebut untuk melakukan klasifikasi menggunakan teknik *Semi-Supervised Learning* berhasil digunakan pada penelitian ini.
- b. Klasifikasi dokumen medis menggunakan teknik *Semi-Supervised Learning* dilakukan dengan membagi percobaan menggunakan porsi data yang berbeda karena dapat menunjukkan nilai akurasi terbaik disetiap porsinya. Setelah model kombinasi dibangun, hasil akurasi yang diperoleh model 4:3 adalah 85%, model 3:4 adalah 87%, model 2:5 adalah 82%, dan model 1:6 adalah 85%.
- c. Hasil akurasi pada proses klasifikasi dokumen medis menggunakan teknik *Semi-Supervised Learning* menguji perbandingan porsi dokumen yang berbeda dan mampu memperoleh akurasi yang sama walaupun perbandingan data yang cukup jauh berbeda porsinya yaitu 4:3 dan 1:6.
- d. Klasifikasi dokumen medis memanfaatkan data tidak berlabel (*unlabeled data*) pada teknik *Semi-Supervised Learning* karena dapat mempengaruhi nilai akurasi sehingga nilai akurasi Sehingga untuk meningkatkan akurasi pemanfaatan data tidak berlabel dapat digunakan dengan jumlah yang bahkan lebih banyak daripada data berlabel.
- e. Untuk menguji model yang sudah dibangun apakah bisa digunakan sebagai acuan untuk melakukan klasifikasi dokumen medis atau tidak, maka model yang dibangun diimplementasi ke dalam sebuah *website* menggunakan *Flask* sebagai *framework*. Pada penelitian ini, *website* berhasil dibangun dan dapat menampilkan halaman *home* yang memuat informasi tabel dan grafik hasil klasifikasi. Selain itu *website* juga dapat

menampilkan halaman *predict* yaitu halaman yang digunakan untuk melakukan prediksi pada teks dokumen medis menggunakan model *classifier* pilihan.

5.2 Saran

Saran yang diberikan untuk pengembangan sistem ini selanjutnya:

- a. Masih terdapat beberapa kata yang tidak baku dari hasil *preprocessing* yang dilakukan.
- b. Hasil klasifikasi pada kategori kecantikan masih banyak yang tidak sesuai dengan *true label* dari kategori, disebabkan karena masalah kecantikan banyak yang sangat mirip dengan masalah kulit. Jadi banyak dari dokumen medis pada kategori kecantikan di prediksi sebagai kategori kulit. Hal tersebut membuat nilai akurasi menurun.
- c. Pengembangan *website* masih perlu menambahkan fitur lain untuk melakukan klasifikasi dengan cara *user* dapat mengupload file dokumen ke dalam sistem.

