

BAB V KESIMPULAN DAN SARAN

5.1. Kesimpulan

Kesimpulan yang diperoleh pada penelitian ini adalah penggunaan fitur ekstraksi dengan menggunakan pustaka *python sklearn* modul *CountVectorizer* dan membandingkan kedua algoritma klasifikasi *Logistic Regression* (LR) dan *Multinomial Naive Bayes* (MNB) dalam mengklasifikasi jenis kelamin berdasarkan nama indonesia, data yang digunakan berjumlah 14908 setelah dilakukannya preprocessing. Penggunaan n-gram dan split data sangat mempengaruhi tingkat akurasi pada algoritma klasifikasi, pengujian dilakukan pada rentang 2 sampai 12 gram dilakukan pencarian nilai rata-ratanya untuk setiap algoritma klasifikasi, dan hasil akurasi mendapatkan skor tertinggi 94.76% pada algoritma *Logistic Regression* (LR) dengan split data 80:20. Hasil validasi model pada *confusion matrix* mendapatkan tingkatan skor yang tidak berbeda jauh pada masing-masing model yang diujikan yaitu 0.94, untuk mengindikasikan bahwa model memiliki tingkat precision dan recall yang baik maka dilihat pada f1-score, untuk nilai f1-score tertinggi didapat pada algoritma *Logistic Regression* (LR) sebesar 0.95. Hasil dari uji prediksi menunjukkan model LR dapat memprediksi lebih baik dibandingkan model MNB dengan tingkat kebenaran prediksi 95.16%. Dapat ditarik kesimpulan, bahwa model *Logistic Regression* (LR) dengan split data 80:20 yang diterapkan ke dalam web aplikasi prediksi jenis kelamin berdasarkan nama.

5.2. Saran

Saran yang dapat diberikan dari penelitian ini untuk pertimbangan penelitian selanjutnya mungkin dapat menggunakan pemanfaatan parameter *analyzer word* pada *CountVectorizer* dengan catatan jika memiliki dataset yang lebih banyak dan data yang sudah lebih bersih (tidak adanya karakter lain selain abjad). Serta dapat melakukan pengujian dengan metode *deep learning*, jika perangkat memenuhi syarat dan perbanyak jumlah data untuk mendapatkan model klasifikasi yang lebih akurat.

Dalam sisi aplikatif bisa ditambahkan fitur *image to text* atau mengubah gambar ke dalam tulisan yang mungkin bermanfaat untuk mendapatkan data keikutsertaan, yang proses pendaftarannya dilakukan secara *offline* dengan cara memfoto daftar keikutsertaan pada kertas.

