

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Sebagai hasil dari evaluasi performa model, ditemukan bahwa penggunaan algoritma Naïve Bayes dalam klasifikasi email spam menunjukkan perubahan signifikan setelah dilakukan penyesuaian pada data. Berikut adalah ringkasan hasil yang diperoleh:

1. Model Naïve Bayes awalnya mencapai akurasi 88%.
2. False Negative (FN) memiliki nilai yang cukup tinggi, yaitu 137, yang berarti banyak email spam yang salah diklasifikasikan sebagai non-spam.
3. False Positive (FP) bernilai 0, menunjukkan bahwa model tidak pernah mengklasifikasikan email non-spam sebagai spam.
4. Ketidakseimbangan dataset, di mana jumlah email non-spam lebih banyak daripada spam, menjadi faktor utama penyebab tingginya nilai False Negative.
5. Untuk mengatasi masalah ini, diterapkan teknik SMOTE guna menyeimbangkan distribusi data.
6. Setelah menerapkan SMOTE, False Negative menurun drastis, dan akurasi model meningkat menjadi 98%.

5.2 Saran

Untuk pengembangan lebih lanjut, beberapa perbaikan dapat diterapkan:

1. Model machine learning dapat dikembangkan menggunakan lebih dari satu algoritma, misalnya kombinasi Naïve Bayes dan SVM untuk perbandingan performa.
2. Jika terdapat ketidakseimbangan kelas seperti jumlah email non-spam yang jauh lebih banyak daripada spam, dapat diterapkan metode penanganan class imbalance lainnya selain SMOTE, seperti undersampling, oversampling, atau cost-sensitive learning.
3. Eksperimen dengan metode ekstraksi fitur yang berbeda, misalnya membandingkan TF-IDF dengan word embeddings (Word2Vec, GloVe, atau FastText) untuk melihat apakah model dapat menangkap pola yang lebih baik dalam data teks.