

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Dalam penelitian ini, empat metode klasifikasi telah dievaluasi untuk mengukur kinerja mereka dalam memprediksi dan mengklasifikasikan data dalam konteks sistem informasi. Metode-metode tersebut meliputi Logistic Regression, Support Vector Machine, Decision Tree, dan K-Nearest Neighbor. Evaluasi dilakukan menggunakan dua skenario pembagian dataset, yaitu rasio 90:10 dan 80:20 untuk data latih dan data uji. Maka dapat disimpulkan seperti dibawah ini:

1. Metode Support Vector Machine (SVM) memiliki kinerja yang konsisten dan baik pada kedua rasio pembagian dataset. SVM mencapai akurasi sekitar 84% (90:10) dan 83% (80:20), serta F-measure sekitar 83% pada kedua rasio tersebut. Ini menunjukkan kemampuan SVM dalam mengklasifikasikan data dengan akurasi tinggi dan keseimbangan antara presisi dan recall.
2. Meskipun metode Logistic Regression memiliki akurasi stabil sekitar 82% pada kedua rasio dataset, F-measure cenderung lebih rendah dibandingkan SVM, terutama pada rasio 80:20.

Dalam konteks penggunaan metode klasifikasi untuk aplikasi dalam teknologi sistem informasi, hasil ini memberikan panduan bahwa metode SVM dan Logistic Regression dapat menjadi pilihan yang baik dalam menghadapi tugas-tugas klasifikasi data. Namun, pemilihan metode yang tepat tetap harus disesuaikan dengan karakteristik dataset dan tujuan akhir dari analisis.

5.2. Saran

Penggunaan dataset saat ini masih mengandung kesalahan, sehingga diharapkan bahwa analisis yang akan datang akan lebih akurat dan optimal. Untuk meningkatkan akurasi pada data teks tweet di Twitter, penerapan metode seperti Support Vector Machine, Logistic Regression, Decision Tree, dan K-Nearest Neighbor dalam meningkatkan akurasi dapat dilakukan dengan mengoptimalkan parameter yang cocok dengan karakteristik dataset. Pendekatan ini diharapkan akan menghasilkan efisiensi dan tingkat akurasi yang maksimal. Namun, jika terjadi masalah missing value pada data yang tidak dapat dihindari, langkah selanjutnya dapat melibatkan pengembangan metode alternatif untuk mengatasi missing value serta menentukan parameter yang sesuai tanpa perlu proses trial and error. Langkah ini diharapkan akan menghasilkan tingkat akurasi yang lebih tinggi pada penelitian selanjutnya.

