

## BAB V PENUTUP

### 5.1 Kesimpulan

Berdasarkan penelitian ini yang menunjukkan perbandingan kinerja algoritma regresi logistik (LR), Naive Bayes (NB), dan Random Forest (RF) dengan pendekatan transformasi dan categorical, berikut adalah kesimpulan yang dapat menjawab rumusan masalah serta selaras dengan tujuan penelitian:

1. Berdasarkan hasil analisis kinerja algoritma Regresi Logistik, Random Forest, dan Naive Bayes untuk klasifikasi data medis penyakit stroke, dapat disimpulkan bahwa algoritma Random Forest menunjukkan performa yang lebih baik dibandingkan dengan algoritma lainnya. Hal ini terlihat dari nilai akurasi, recall, precision, dan F1-score yang lebih tinggi pada berbagai skenario pembagian data. Regresi Logistik dengan transformasi memiliki kinerja yang lebih baik dibandingkan dengan tanpa transformasi, sedangkan Naive Bayes menunjukkan hasil yang lebih rendah dibandingkan dua algoritma lainnya. Dengan demikian, pemilihan algoritma dalam klasifikasi data medis penyakit stroke harus mempertimbangkan karakteristik data serta kebutuhan spesifik dari proses klasifikasi.
2. Dalam membandingkan kinerja ketiga algoritma ini, terlihat bahwa Random Forest dengan transformasi consistently memiliki nilai akurasi tertinggi dalam berbagai skenario pembagian data. Pada pembagian data (90:10), (80:20), dan (70:30), Random Forest dengan transformasi mencapai akurasi sebesar 94%, sedangkan Random Forest dengan categorical mencapai akurasi 85%. Algoritma Regresi Logistik dan Naive Bayes memiliki nilai akurasi yang lebih rendah, dengan nilai tertinggi pada Regresi Logistik dengan transformasi yang mencapai 82% pada pembagian data (90:10). Hal ini menunjukkan bahwa dalam konteks klasifikasi data medis penyakit stroke, algoritma Random Forest lebih efektif dibandingkan Regresi Logistik dan Naive Bayes.

3. Analisis terhadap keunggulan dan kelemahan masing-masing algoritma menunjukkan bahwa Random Forest memiliki keunggulan dalam semua metrik evaluasi, termasuk recall dan precision yang tinggi. Sementara itu, Regresi Logistik dengan transformasi memiliki keunggulan dibandingkan dengan tanpa transformasi, tetapi masih kalah dibandingkan Random Forest. Naive Bayes menunjukkan performa yang lebih rendah, terutama pada recall yang hanya mencapai 49% dalam beberapa skenario, yang menunjukkan bahwa model ini kurang mampu mengidentifikasi kasus positif dengan baik. Oleh karena itu, dalam pemilihan algoritma untuk klasifikasi data medis penyakit stroke, Random Forest dapat menjadi pilihan utama, sementara Regresi Logistik dapat digunakan sebagai alternatif, sedangkan Naive Bayes kurang direkomendasikan.
4. Dari hasil uji coba matriks evaluasi terhadap diagnosis penyakit berdasarkan data medis stroke, terlihat bahwa algoritma Random Forest memberikan hasil yang lebih optimal dibandingkan dengan algoritma lainnya. Nilai F1-score yang diperoleh oleh Random Forest dengan transformasi mencapai 93% dalam berbagai pembagian data, menunjukkan keseimbangan antara precision dan recall. Regresi Logistik dengan transformasi memiliki F1-score maksimal sebesar 81%, sementara Naive Bayes hanya mencapai maksimal 78%. Dengan demikian, dalam penerapan diagnosis penyakit stroke berbasis data medis, penggunaan algoritma Random Forest lebih direkomendasikan untuk meningkatkan akurasi klasifikasi dan mengurangi kesalahan dalam diagnosis.

Secara keseluruhan, algoritma Random Forest dengan transformasi data memberikan hasil terbaik dalam semua metrik evaluasi, terutama dalam hal akurasi, recall, dan F1-Score. Regresi Logistik dan Naive Bayes memiliki performa yang lebih rendah dibandingkan Random Forest, namun keduanya masih memiliki aplikasi potensial tergantung pada karakteristik dataset yang digunakan. Dengan demikian, penggunaan Random Forest lebih direkomendasikan untuk klasifikasi data medis dalam penelitian ini.

## 5.2 Saran

Berdasarkan hasil penelitian ini, beberapa saran dapat diajukan untuk penelitian selanjutnya guna memperdalam analisis dan mengembangkan metodologi klasifikasi data medis.

1. penelitian selanjutnya dapat mengeksplorasi penggunaan algoritma klasifikasi lain, untuk melihat apakah algoritma tersebut dapat memberikan hasil yang lebih baik dibandingkan algoritma yang digunakan dalam penelitian ini. Dengan mengeksplorasi lebih banyak algoritma, dapat diidentifikasi metode yang paling optimal untuk klasifikasi data medis.
2. Penelitian selanjutnya disarankan untuk menerapkan teknik praproses data yang lebih beragam, seperti deteksi dan penanganan outlier menggunakan metode Z-score atau IQR, transformasi data seperti Principal Component Analysis (PCA) untuk reduksi dimensi. Selain itu, pendekatan ensemble learning yang menggabungkan beberapa algoritma, seperti Stacking atau AdaBoost, juga dapat dieksplorasi untuk meningkatkan akurasi dan performa model klasifikasi.
3. Implementasi Sistem Prediksi Stroke Berbasis Pembelajaran Mesin Puskesmas dan instansi kesehatan dapat mempertimbangkan untuk mengimplementasikan sistem berbasis pembelajaran mesin dalam proses screening awal penyakit stroke. Dengan menerapkan algoritma seperti Random Forest yang telah terbukti memiliki performa tinggi dalam klasifikasi data medis, tenaga kesehatan dapat memperoleh prediksi yang lebih cepat dan akurat. Sistem ini dapat digunakan sebagai alat bantu dalam mengidentifikasi pasien dengan risiko tinggi, sehingga tindakan pencegahan dan penanganan dapat dilakukan lebih dini. Implementasi teknologi ini juga dapat membantu efisiensi kerja tenaga kesehatan dalam menangani jumlah pasien yang besar.
4. Kolaborasi dengan Institusi Akademik dan Peneliti Puskesmas dan instansi kesehatan dapat menjalin kerja sama dengan institusi akademik dan peneliti dalam mengembangkan model prediksi yang lebih akurat dan sesuai dengan karakteristik populasi setempat. Kolaborasi ini dapat mencakup pengumpulan data medis yang lebih komprehensif,

pengujian algoritma baru, serta evaluasi efektivitas sistem prediksi dalam lingkungan medis yang nyata. Selain itu, kerja sama dengan akademisi juga dapat membuka peluang untuk penelitian lebih lanjut dalam bidang kesehatan berbasis data.

