

## BAB 5 SIMPULAN DAN SARAN

### 5.1 Simpulan

Berdasarkan penelitian yang sudah dilakukan, implementasi nilai *information gain* untuk melakukan *feature selection* dengan menggunakan nilai akhir *threshold* median dapat diimplementasikan. Penelitian ini menggunakan 8 dataset yang memiliki nilai *features* lebih dari 50, yang didapatkan dari *website* UCI Machine Learning Repository dan Open ML. Dengan mengimplementasi *feature selection* dengan nilai *threshold* median pada dataset yang digunakan, dan dilakukan pengujian klasifikasi menggunakan 4 algoritma, yaitu *logistic regression*, *decision tree*, *random forest*, *naive bayes*. Hasil dari pengujian tersebut menunjukkan bahwa rata-rata nilai *accuracy*, *precision*, *recall*, dan *f1-score* lebih besar dari 51%. Dataset yang digunakan dalam penelitian ini meliputi: AP\_Breast\_Omentum, Musk (Version 2), Internet Advertisements, Bioresponse, Arcene, AP\_Colon\_Kidney, Hill Valley, dan Nomao.

Penggunaan algoritma *logistic regression* pada berbagai dataset telah memberikan hasil evaluasi bahwa nilai *accuracy* yang didapatkan untuk semua dataset lebih dari 76%. Pada dataset AP\_Breast\_Omentum, *accuracy* adalah 97%. Untuk dataset Musk (Version 2), *accuracy* mencapai 99%. Arcene menghasilkan *accuracy* 90%. Sementara itu, dataset Internet Advertisement mencapai *accuracy* 96%. Dataset Bioresponse memberikan *accuracy* 76%. AP\_Colon\_Kidney menghasilkan *accuracy* 98%. Hill Valley mencapai *accuracy* 90%. Terakhir, dataset Nomao mencapai *accuracy* 82%.

Penggunaan algoritma *decision tree* pada berbagai dataset telah memberikan hasil evaluasi bahwa nilai *accuracy* yang didapatkan untuk semua dataset lebih dari 56%. Pada dataset AP\_Breast\_Omentum, *accuracy* adalah 95%. Untuk dataset Musk (Version 2), *accuracy* mencapai 95%. Arcene menghasilkan *accuracy* 73%. Sementara itu, dataset Internet Advertisement mencapai *accuracy* 95%. Dataset Bioresponse memberikan *accuracy* 72%. AP\_Colon\_Kidney menghasilkan *accuracy* 96%. Hill Valley mencapai *accuracy* 56%. Terakhir, dataset Nomao mencapai *accuracy* 90%.

Penggunaan algoritma *random forest* pada berbagai dataset telah memberikan hasil evaluasi bahwa nilai *accuracy* yang didapatkan untuk semua

dataset lebih dari 56%. Pada dataset AP\_Breast\_Omentum, *accuracy* adalah 96%. Untuk dataset Musk (Version 2), *accuracy* mencapai 92%. Arcene menghasilkan *accuracy* 85%. Sementara itu, dataset Internet Advertisement mencapai *accuracy* 97%. Dataset Bioresponse memberikan *accuracy* 80%. AP\_Colon\_Kidney menghasilkan *accuracy* 97%. Hill Valley mencapai *accuracy* 56%. Terakhir, dataset Nomao mencapai *accuracy* 94%.

Penggunaan algoritma *naive bayes* pada berbagai dataset telah memberikan hasil evaluasi bahwa nilai *accuracy* yang didapatkan untuk semua dataset lebih dari 51%. Pada dataset AP\_Breast\_Omentum, *accuracy* adalah 96%. Untuk dataset Musk (Version 2), *accuracy* mencapai 84%. Arcene menghasilkan *accuracy* 67%. Sementara itu, dataset Internet Advertisement mencapai *accuracy* 93%. Dataset Bioresponse memberikan *accuracy* 63%. AP\_Colon\_Kidney menghasilkan *accuracy* 97%. Hill Valley mencapai *accuracy* 51%. Terakhir, dataset Nomao mencapai *accuracy* 81%.

Berdasarkan hasil pengujian, dapat disimpulkan bahwa algoritma *logistic regression* menunjukkan performa unggul dalam klasifikasi menggunakan dataset hasil *feature selection*, dengan memiliki nilai *accuracy* yang lebih baik dibandingkan dengan algoritma lainnya. Lebih khususnya, dari 8 dataset yang digunakan, algoritma *logistic regression* unggul pada 5 dataset, yaitu: AP\_Breast\_Omentum, Musk (Version 2), Arcene, Hill Valley, dan AP\_Colon\_Kidney, sementara algoritma *decision tree* tidak unggul pada satu pun dataset, *random forest* unggul pada 3 dataset, dan *naive bayes* tidak unggul pada satu pun dataset.

## 5.2 Saran

Berdasarkan hasil penelitian, terdapat beberapa saran untuk proses penelitian di masa depan.

1. Melakukan proses *feature selection* dengan nilai *threshold* mean, dan menggunakan metode *backward feature elimination*.
2. Melakukan uji coba model dengan menggunakan algoritma yang berbeda, seperti contoh *neural network* atau *support vector machine*.