

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Penelitian ini bertujuan untuk menangani distribusi kelas yang tidak seimbang pada dataset menggunakan teknik resampling yaitu over-sampling. Pengujian perbandingan teknik resampling telah dilakukan dengan membandingkan antara penerapan teknik resampling dengan dataset yang tidak menggunakan teknik resampling. Teknik resampling Synthetic Minority Over-Sampling Technique (SMOTE) digunakan untuk menangani ketidakseimbangan kelas pada dataset. Berdasarkan hasil percobaan maupun pengujian yang telah dilakukan dapat ditarik kesimpulan :

1. Penanganan distribusi kelas yang tidak seimbang pada dataset menggunakan algoritma SMOTE dapat meningkatkan nilai akurasi maupun g-mean pada algoritma Naïve Bayes. Hal tersebut menunjukkan bahwa proses penanganan terhadap distribusi kelas yang tidak seimbang pada tahap pra-pemrosesan data memberikan pengaruh terhadap nilai akurasi maupun g-mean algoritma Naïve Bayes. Pada scenario percobaan yang telah dilakukan algoritma Naïve Bayes memiliki akurasi paling baik 96,43 % pada dataset ecoli 15,8 setelah dilakukan SMOTE dengan 10 fold cross validation, dan memiliki nilai G-mean paling baik 96,42 % pada dataset ecoli 15,8 setelah dilakukan SMOTE dengan 10 fold cross validation.
2. Penanganan distribusi kelas yang tidak seimbang pada dataset menggunakan algoritma SMOTE dapat meningkatkan nilai akurasi maupun

g-mean pada algoritma SVM. Hal tersebut menunjukkan bahwa proses penanganan terhadap distribusi kelas yang tidak seimbang pada tahap pra-pemrosesan data memberikan pengaruh terhadap nilai akurasi maupun g-mean algoritma SVM. Pada scenario percobaan yang telah dilakukan algoritma SVM memiliki akurasi paling baik 99,02 % pada dataset ecoli 15,8 setelah dilakukan SMOTE dengan 10 fold cross validation, dan memiliki nilai G-mean paling baik 99,37 % pada dataset ecoli 15,8 sebelum dilakukan SMOTE dengan 10 fold cross validation.

3. Penanganan distribusi kelas yang tidak seimbang pada dataset menggunakan algoritma SMOTE dapat meningkatkan nilai akurasi maupun g-mean pada algoritma KNN. Hal tersebut menunjukkan bahwa proses penanganan terhadap distribusi kelas yang tidak seimbang pada tahap pra-pemrosesan data memberikan pengaruh terhadap nilai akurasi maupun g-mean algoritme KNN. Pada scenario percobaan yang telah dilakukan algoritma SVM memiliki akurasi paling baik 97,29 % pada dataset ecoli 15,8 setelah dilakukan SMOTE dengan 10 fold cross validation, dan memiliki nilai G-mean paling baik 99,53 % pada dataset ecoli 15,8 sebelum dilakukan SMOTE dengan 10 fold cross validation.
4. Penanganan distribusi kelas yang tidak seimbang pada dataset menggunakan algoritme SMOTE dapat meningkatkan nilai akurasi maupun g-mean pada algoritma Decision Tree. Hal tersebut menunjukkan bahwa proses penanganan terhadap distribusi kelas yang tidak seimbang pada tahap

pra-pemrosesan data memberikan pengaruh terhadap nilai akurasi maupun g-mean algoritme Decision Tree. Pada scenario percobaan yang telah dilakukan algoritma SVM memiliki akurasi paling baik 97,29 % pada dataset ecoli 15,8 sebelum dilakukan SMOTE dengan 80% data training, dan memiliki nilai G-mean paling baik 96,29 % pada dataset ecoli 15,8 setelah dilakukan SMOTE dengan 10 fold cross validation.

5. Nilai akurasi dan nilai G-mean algoritma Naïve Bayes konsisten dengan performanya pada setiap level imbalance ratio, sebelum implementasi SMOTE memiliki performa yang tidak baik, sedangkan setelah diimplementasikan SMOTE algoritma Naïve Bayes memiliki peningkatan akurasi yang konsisten. Sehingga dapat ditarik kesimpulan bahwa kombinasi SMOTE + Naïve Bayes paling efektif digunakan pada dataset imbalance dengan level yang berbeda-beda pada skema 10 fold cross validation maupun 80% data testing yang diujikan sebanyak 50 kali.
6. Implementasi dari algoritma SMOTE membantu dalam pengklasifikasian terhadap data kelas minoritas yang menjadi masalah pada latarbelakang penelitian ini, hal tersebut dapat dilihat pada tabel confusion matrix bagian nilai FN atau false negative dimana jika dibandingkan dengan tabel confusion matrix sebelum dan sesudah implementasi SMOTE terdapat penurunan dari nilai FN atau false negative. Selain itu dapat dilihat pada perbandingan nilai g-mean sebelum implementasi SMOTE dan setelah implementasi SMOTE.

## 5.2 Saran

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa saran untuk penelitian selanjutnya, yakni :

1. Untuk penelitian selanjutnya dapat dikembangkan dengan mengombinasikan teknik resampling antara oversampling dan undersampling.
2. Pada penelitian selanjutnya dapat dibandingkan hasil penelitian yang telah dilakukan ini dengan metode ensemble yaitu pendekatan pada level algoritma.
3. Untuk penelitian kedepannya bisa kombinasi antara pendekatan pada level data dan pendekatan pada level algoritma untuk mendapatkan performa klasifikasi pada data imbalance yang baik.

