

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini memiliki tujuan untuk menangani distribusi kelas tidak seimbang pada dataset menggunakan metode ensemble *Balanced-Bagging* yang diaplikasikan pada algoritma klasifikasi *Decision Tree*. Berdasarkan pada penelitian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

1. Implementasi metode *ensemble Balanced-Bagging* dan *Bagging* terhadap algoritma klasifikasi mampu meningkatkan kinerja hasil akurasi (*balanced accuracy*) dan nilai *geometric mean*, namun pada beberapa kasus didapati hasil evaluasi dari *bagging* lebih rendah daripada hasil evaluasi dari algoritma klasifikasi, kondisi ini bergantung pada karakteristik data dan keberadaan *noise sample* dalam setiap dataset.
2. Teknik *Stratified K-fold Cross Validation* pada proses *splitting data* mampu meningkatkan hasil evaluasi pada beberapa dataset jika dibandingkan dengan teknik *K-Fold* tanpa stratifikasi terlebih dahulu.
3. Pemanfaatan teknik *undersampling* pada metode *bagging* mampu meningkatkan hasil evaluasi dari metode *bagging* secara signifikan.
4. Metode *bagging* mengalami peningkatan saat menghadapi data yang *overfitting*, dan saat dihadapkan pada data yang kurang berisik (*noise*) maka *bagging* akan mengalami penurunan hasil kinerja dan berdampak pula pada nilai dari hasil evaluasi.
5. Sementara dalam hal kinerja metode *balanced-bagging* mampu bekerja dengan baik pada setiap dataset sehingga hasil akurasi dan *geometric mean* dari metode *balanced-bagging* juga baik.
6. Saat dilakukan perbandingan antara metode *bagging* dan *balanced-bagging* terhadap hasil akurasi dan *geometric mean*, menunjukkan bahwa hasil kinerja

dari metode balanced-bagging mampu melebihi metode bagging pada keseluruhan dataset.

5.2 Saran

Berdasarkan pada penelitian yang sudah dilakukan terdapat saran untuk penelitian selanjutnya terkait dengan masalah dataset imbalance:

1. Penelitian selanjutnya dapat diujikan dengan menggunakan dataset multiclass.
2. Penelitian selanjutnya dapat digunakan dataset sejenis dengan variasi imbalance ratio.
3. Dataset yang diujikan dapat berupa *data categorical* maupun *data continuous*.
4. Dataset yang digunakan dapat diujikan pada metode Boosting atau Stacking.
5. Pada penelitian selanjutnya dapat dilakukan kombinasi antara metode *oversampling* dan *bagging*.