

BAB VI KESIMPULAN DAN SARAN

Dalam bab ini akan dijabarkan kesimpulan yang dapat diambil dari hasil evaluasi pada bab V dan saran yang diberikan untuk pengembangan penelitian berikutnya.

6.1. Kesimpulan

Berdasarkan evaluasi yang dilakukan dalam penelitian tugas akhir ini, maka dapat ditarik kesimpulan sebagai berikut:

1. *Dataset training* untuk *low imbalanced dataset* mengusulkan metode KSMOTE-SVM sebagai usulan utama, dan metode SSO-SVM sebagai usulan kedua, berdasarkan hasil evaluasi *low imbalanced dataset* pada subbab 5.1.1.1. Usulan didapatkan dari hasil persentase *correctly classified instances* tertinggi, dan hasil persentase *confusion matrix* yang berhasil mengklasifikasi semua *class* (*class* mayoritas dan *class* minoritas) dan persentase *misclassified* terkecil.
2. *Dataset training* untuk *medium imbalanced dataset* berdasarkan hasil evaluasi *medium imbalanced dataset* pada subbab 5.1.1.2 dengan hasil persentase *correctly classified instances* tertinggi, dan hasil persentase *confusion matrix* yang berhasil mengklasifikasi semua *class* (*class* mayoritas dan *class* minoritas) dan persentase *misclassified* terkecil mengusulkan metode KSMOTE-SVM sebagai usulan utama, dan metode SSO-SVM sebagai usulan kedua.
3. *Dataset training* untuk *high imbalanced dataset* berdasarkan hasil evaluasi *high imbalanced dataset* pada subbab 5.1.1.3 mengusulkan metode AdaBoostSVM sebagai usulan utama, dan metode KSMOTE-SVM sebagai usulan kedua, dengan melihat hasil persentase *correctly classified instances* tertinggi, dan hasil persentase *confusion matrix* yang berhasil mengklasifikasi semua *label class* (*class* mayoritas dan *class* minoritas) dan persentase *misclassified* terkecil.

4. *Dataset* Mahasiswa dengan kategori *medium imbalanced dataset* dapat diklasifikasi dengan metode usulan utama yaitu, metode KSMOTE-SVM dan usulan kedua yaitu, metode SSO-SVM. Hasil evaluasi pada subbab 5.1.2.1, 5.1.2.2, dan 5.1.2.4 menunjukkan KSMOTE-SVM sebagai metode yang disarankan berdasarkan hasil persentase *correctly classified instances* tertinggi dan persentase *misclassified* terkecil dengan *label class* minoritas dapat terklasifikasi, dan
5. *Dataset* Mahasiswa dengan kategori *high imbalanced dataset* berdasarkan hasil evaluasi pada subbab 5.1.2.3 dan 5.1.2.5 menunjukkan *dataset* hanya dapat diklasifikasikan dengan metode usulan kedua yaitu, metode KSMOTE-SVM, karena metode selain KSMOTE-SVM mengalami salah klasifikasi (*misclassified*) pada *label class* minoritas menyebabkan metode KSMOTE-SVM memiliki persentase *misclassified* terkecil. KSMOTE-SVM juga memiliki hasil persentase *correctly classified instances* tertinggi.
6. Metode KSMOTE-SVM atau oversampling, hanya bekerja dengan baik pada *dataset* yang memiliki dua *class* karena metode KSMOTE-SVM hanya melakukan *oversampling* pada satu *class* minoritas. Jika *dataset* memiliki *multiclass* maka tidak semua *class* dapat di-*oversampling*, tetapi *oversampling* pada *class* yang paling minoritas. Metode KSMOTE-SVM cocok digunakan pada semua karakteristik data *imbalanced* (*low*, *medium*, dan *high imbalanced dataset*) sesuai dengan hasil keseluruhan evaluasi data *training* (subbab 5.1.1.4) dan data *testing* (subbab 5.1.2.6).
7. Metode AdaBoostSVM merupakan metode untuk memperbaiki akurasi SVM sehingga dapat bekerja pada *dataset* dengan dua *class* maupun *multiclass*. Metode AdaboostSVM tidak dapat memberikan hasil yang baik jika metode SVM mengalami *missclassified* pada salah satu *class*, karena metode AdaBoostSVM tidak dapat memperbaiki akurasi pada *class* yang sama sekali *missclassified*, hal ini dapat terlihat pada hasil evaluasi pada subbab V dengan Tabel 5.5, Tabel 5.8, Tabel 5.26, Tabel 5.32.

8. Metode SSO-SVM atau *undersampling*, hanya melakukan *undersampling* pada satu *class* paling mayoritas, sehingga membuat metode ini bekerja dengan efektif pada *dataset* yang memiliki dua *class*. Jika metode digunakan pada *dataset multiclass* metode tidak dapat menyeimbangkan semua *class* yang terdapat pada *dataset*. SSO-SVM dapat bekerja pada karakteristik data *low imbalanced dataset* dan *medium imbalanced dataset* sesuai dengan hasil keseluruhan evaluasi data *training* dengan karakteristik *low* dan *medium imbalanced dataset* (subbab 5.1.1.4) dan data *testing* dengan karakteristik *medium imbalanced dataset* (subbab 5.1.2.6). Pada hasil evaluasi data *training* dalam subbab 5.1.1.3 menunjukkan metode SSO-SVM tidak bekerja dengan baik pada *high imbalanced dataset*.
9. Jika melakukan analisis pada *dataset* baru, maka sebaiknya diperhatikan jumlah *class* yang dimiliki oleh *dataset* dan *imbalance ratio* pada *class* dalam *dataset*. Jumlah *class* mempengaruhi metode yang digunakan dikarenakan terdapat metode yang hanya dapat bekerja pada *dataset* dengan dua *class*, sehingga tidak efektif jika diterapkan pada *multiclass dataset*. *Imbalance ratio* yang dimiliki oleh *dataset* dapat mempengaruhi hasil evaluasi, jika menggunakan *classifier* menemukan *class* yang seluruhnya *missclassified* disarankan untuk menggunakan metode yang menerapkan *oversampling* atau *undersampling*, karena optimasi *classifier* seperti *boosting* tidak efektif dalam menangani permasalahan *missclassified class* pada *imbalanced dataset*.

6.2. Saran

Adapun saran-saran yang diperoleh untuk pengembangan penelitian tugas akhir ini, yaitu:

1. Menggunakan metode klasifikasi lain yang dapat di kombinasikan dengan *support vector machine* untuk menyelesaikan permasalahan *imbalanced data* selain yang digunakan pada tugas akhir,
2. Melakukan eksperimen dan analisis terhadap *dataset* dengan tipe data *numeric (integer dan real)*, dan
3. Memperlihatkan bentuk visual dari hyperplane yang tercipta oleh *support vector machine*.