

BAB 5

SIMPULAN DAN SARAN

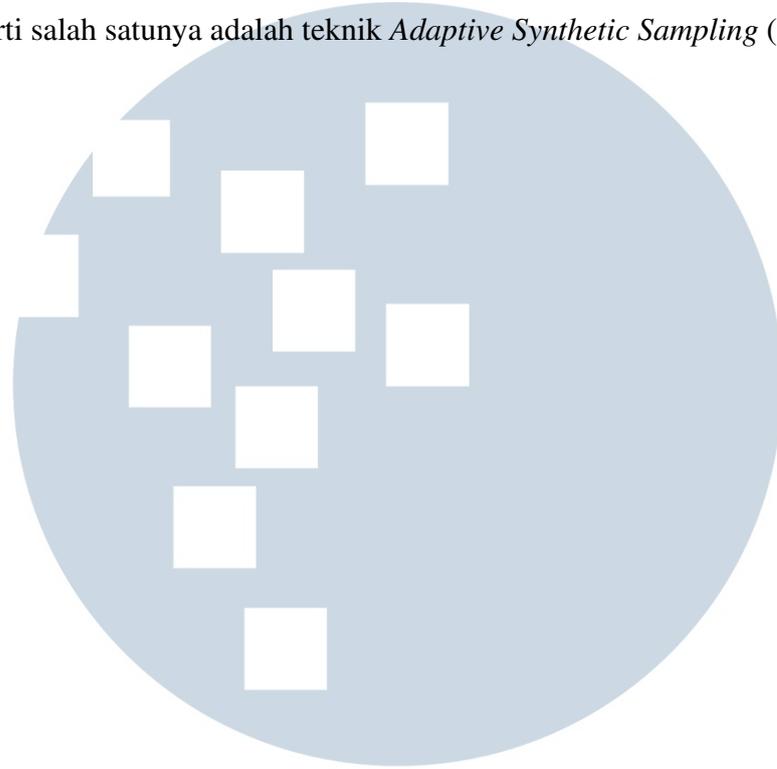
5.1 Simpulan

1. Penelitian analisa sentimen pengguna media sosial twitter terhadap minat membayar pajak menggunakan algoritma *naive bayes* telah berhasil dilakukan dengan membandingkan tiga variasi *naive bayes*. Data yang digunakan pada penelitian ini merupakan data hasil *crawling data* twitter menggunakan bantuan dari dua *scrapping tools*, yaitu *tweet-harvest* dan *snsrape* yang kemudian diberi label secara manual oleh tiga orang dan diambil label akhir berdasarkan suara label terbanyak. Data yang didapatkan dan dilakukan *pre-processing* berjumlah 2617 data dengan label akhir sebanyak 295 data berlabel positif, 145 data berlabel netral, dan 2177 data berlabel negatif.
2. Berdasarkan hasil uji coba yang telah dilakukan terhadap data yang tidak seimbang menggunakan evaluasi *confusion matrix*, implementasi *bernoulli naive bayes* menggunakan teknik *Synthetic Minority Oversampling Technique* (SMOTE) pada perbandingan dataset 80:20 menghasilkan performa yang lebih baik dibandingkan variasi *gaussian* dan *multinomial naive bayes* dari ketiga skenario pembagian dataset. Hal ini didapatkan dari hasil performa yang dihasilkan yang memiliki nilai akurasi sebesar 91.03%, *f1-score* 71.18%, *precision* 71.11%, dan *recall* sebesar 71.43%. Pembagian data latihan (*training*) dan data uji (*testing*) dalam tiga skenario uji coba menunjukkan bahwa semakin banyak data latihan (*training*) yang digunakan maka performa yang dihasilkan akan semakin tinggi. Selain itu, penggunaan teknik SMOTE terhadap data yang tidak seimbang dapat membantu meningkatkan *multinomial naive bayes* dan *bernoulli naive bayes*. Namun, tidak dapat meningkatkan hasil performa dari *gaussian naive bayes*.

5.2 Saran

1. Mengumpulkan lebih banyak data dengan menambahkan kata kunci baru dengan tujuan untuk menyeimbangkan data dan membuat pemodelan yang telah digunakan dapat menganalisa sentimen kata yang lebih bervariasi.

2. Mencoba menggunakan teknik *resampling* yang lainnya guna menyeimbangkan perbandingan kelas pada dataset yang tidak seimbang, seperti salah satunya adalah teknik *Adaptive Synthetic Sampling* (ADASYN)



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA