

## **BAB 5 PENUTUP**

### **5.1 Kesimpulan**

Setelah dilakukan penelitian terhadap prediksi serangan dengan menggunakan *Naïve Bayes* dan membandingkannya dengan penelitian sebelumnya yang menggunakan algoritma *Hidden Markov Model* dengan data yang sama, sehingga dapat disimpulkan sebagai berikut:

1. Hasil evaluasi yang didapat dengan melakukan percobaan sebanyak tiga skenario dengan algoritma *Naïve Bayes* ialah:
  - Skenario 1 dengan pembagian data *training* 70% dan *testing* 30% menghasilkan nilai akurasi sebesar 75,68%, nilai presisi sebesar 67,56%, dan nilai *recall* sebesar 98,98%.
  - Skenario 2 dengan pembagian data *training* 80% dan *testing* 20% menghasilkan nilai akurasi sebesar 75,01%, nilai presisi sebesar 67,48%, dan nilai *recall* sebesar 99,01%.
  - Skenario 3 dengan pembagian data *training* 90% dan *testing* 10% menghasilkan nilai akurasi sebesar 75,63%, nilai presisi sebesar 67,52%, dan nilai *recall* sebesar 98,98%.
2. Metode *Oversampling* SMOTE berhasil dijalankan sehingga mendapatkan nilai *probability* yang seimbang tiap *class* tetapi hasil akhir lainnya dipengaruhi oleh nilai tiap label pada masing-masing baris data.
3. Algoritma *Naïve Bayes* memiliki nilai akurasi yang hampir serupa dengan algoritma *Hidden Markov Model* dalam memprediksi sebuah serangan pada *data log web server* UPNVJ. Nilai presisi algoritma *Naïve Bayes* lebih rendah sekitar 20-30% dibanding algoritma *Hidden Markov Model* tetapi algoritma *Naïve Bayes* memiliki nilai *recall* yang cukup tinggi dibandingkan dengan algoritma *Hidden Markov Model*.

## 5.2 Saran

Berdasarkan penelitian yang telah dilakukan, diperlukannya suatu pengujian dan pengembangan penelitian lebih lanjut mengenai data dan hasil prediksi dengan menggunakan agar menjadi lebih baik dan sistem lebih tepat dalam memprediksi. Saran untuk penelitian ini ialah sebagai berikut.

1. Penggunaan jenis *Naïve Bayes* yang lain (*Gaussian Naïve Bayes* atau *Multinomial Naïve Bayes*) untuk pengolahan dataset yang digunakan pada penelitian ini.
2. Adanya penambahan label dalam data agar lebih banyak varian pada jenis label yang digunakan untuk melakukan *training* pada algoritma *Naïve Bayes* sehingga penggunaan model probabilitas bisa lebih baik.
3. Menghilangkan *data noise* pada data yang bernilai (0) pada tiap-tiap label yang digunakan.
4. Perbandingan dengan algoritma *Data Mining* lain seperti *Support Vector Machine* untuk menguji dataset yang digunakan dalam penelitian ini.
5. Pengujian menggunakan metode *undersampling*.