

## BAB 5

### PENUTUP

#### 5.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka dapat diambil kesimpulan sebagai berikut:

1. Model menggunakan algoritma *Random Forest* untuk pendeteksian situs web *phishing* dibangun menggunakan *dataset* dari *UCI Machine Learning Repository* dengan data berjumlah 11055 yang terdiri dari 6157 situs web *phishing* dan 4898 situs web asli. *Dataset* mempunyai 30 fitur yang mana dalam proses pembangunan model dilakukan seleksi fitur berdasarkan apakah fitur tersebut membutuhkan akses melalui pihak ketiga atau melalui *web service* saat ekstraksi fitur, hal ini dilakukan agar dapat meningkatkan *usability*-nya yaitu membuat pendeteksian bisa dilakukan dengan lebih cepat. Jadi jumlah fitur yang digunakan setelah diseleksi berjumlah 16 fitur. Sebelum dilakukan pelatihan data dibagi menjadi 70% untuk data latih dan 30% untuk data uji. Untuk menentukan parameter terbaik dan melakukan *cross-validation*, digunakan *grid search* dengan jumlah *cross-validation* 5 kali. Didapatkan dari melalui *tuning* parameter bahwa jumlah pohon optimal untuk *dataset* ini adalah 80 pohon. Proses pembangunan model seluruhnya menggunakan *library* Scikit-Learn pada Python. Evaluasi model dilakukan menggunakan *confusion matrix* yang didapatkan hasil akurasi 88%, *recall* 84% dan presisi 91,3%.
2. Model yang sudah dibangun diimplementasikan ke ekstensi peramban Google Chrome dengan mengubahnya ke bentuk format JSON. Atribut-atribut yang ada diambil dari model menggunakan bantuan *library* tree pada Sklearn dan JSON untuk men-*dump file*. Untuk ekstraksi fitur dari halaman situs web diambil dengan fungsi-fungsi yang mempunyai aturan yang mengembalikan nilai  $\{-1, 0, 1\}$ . Klasifikasi atau prediksi dilakukan

dengan mengambil *voting* terbanyak dari hasil setiap prediksi *Decision Tree* pada *Random Forest*.

3. Pengujian ekstensi Google Chrome dilakukan menggunakan 100 data baru yang berisi 50 situs web *phishing* dan 50 situs web asli. Hasil kinerja dari ekstensi yang dibuat menggunakan model algoritma *Random Forest* mempunyai akurasi 88%, *recall* 84% dan presisi 91,3%.
4. Penelitian ini bisa dikatakan cukup berhasil bila dibandingkan kinerjanya dengan deteksi situs web *phishing* dengan *proteksi* default dari *Google Safe Browsing* (GSB) pada Google Chrome yang mempunyai rata-rata akurasinya yaitu ~45% saja.

## 5.2. Saran

Berdasarkan penelitian yang telah dilakukan maka penulis menyarankan beberapa hal sebagai berikut:

1. Menggunakan *dataset* yang mempunyai fitur dan jumlah data yang lebih banyak agar hasil klasifikasi dapat lebih akurat.
2. Mencoba algoritma *machine learning* lainnya yang bisa dipadukan dengan pengambilan ekstraksi fitur seperti contoh menggunakan citra dari tampilan suatu situs web agar diketahui ciri suatu situs *web phishing* tidak hanya dari URL atau DOM saja.