BAB I
PENDAHULUAN

1.1 Latar Belakang

Website merupakan salah satu aplikasi populer bagi pengguna internet yang bersifat publik. Oleh karena itu, *website* menjadi salah satu pilihan bagi *user* dalam membantu pekerjaannya sehari-hari dengan penggunaan yang mudah serta bisa diakses dimanapun dan kapanpun. Namun, dikarenakan sifatnya yang publik *website* sering mengalami serangan. Sehingga, menyebabkan suatu kerusakan pada *web server* yang mengelolah *website* tersebut. *Web server* yang memiliki tingkat keamanan yang lemah selalu menjadi sasaran yang tepat bagi para *attacker* saat menyerang *web server*.

Ancaman-ancaman pada *website* yang terjadi pada tahun 2017 sudah didata oleh OWASP (*Open Web Application Security Project*) dan sudah tercatat pada OWASP Top 10 Security – 2017. Tingkat ancaman yang diberi nilai sudah dihitung dengan kalkulator khusus dari NIST (*National Institute of Standards and Technology*) yang disebut CVSS (*Common Vunlnerability Scoring System*). Namun, segala kegiatan pada *web server* khususnya kegiatan penyerangan terhadap *website* telah dicatat pada *log* dari *web server*, macam-macam *log* yang tercatat ialah *access log* dan *error log.*

Pada tahun 2020, Rico Andreas melakukan penelitian yang berjudul Memprediksi Serangan Pada SIM (*Security Information Management*) Dengan Menggunakan Algoritma *Hidden Markov Model* telah dilakukan prediksi dan klasifikasi pada dataset sebuah serangan yang ada pada *log* dengan menggunakan algoritma *Hidden Markov Model.* Dengan hasil evaluasi akhir yang dihasilkan terbilang tidak begitu baik karena dari tiga skenario yang dijalankan peneliti sebelumnya, akurasi yang dihasilkan berada dibawah 80%, presisi tertinggi yang dihasilkan hanya 90% serta recall yang berada dibawah 15%. Lalu pada prediksi sebuah kelas serangan *(class attack)* algoritma *Hidden Markov Model* mendapatkan hasil prediksi sebesar 85%, keakuratan dalam memprediksi serangan dan bukan serangan sebesar 70% dan dari banyaknya kelas serangan

pada data, model hanya dapat memprediksi benar sebesar 15%.

Pada penelitian kali ini, akan dilakukan percobaan mengukur tingkat akurasi, presisi dan *recall* dengan menggunakan dataset yang sama dengan menggunakan variabel yang sama pula agar penelitian ini bisa *apple-to-apple* dengan penelitian sebelumnya.

1.2 Rumusan Masalah

Berdasarkan dari latar belakang yang sudah dijelaskan, maka dapat diangkat sebuah rumusan masalah yaitu.

1. Apakah *Naïve Bayes* dapat memprediksi lebih baik dibanding *Hidden Markov Model* pada sebuah serangan atau tidak pada *log* Apache?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini ialah.

1. Untuk melihat prediksi serangan atau tidak serangan yang terjadi pada *website* pada data *log* Apache.
2. Untuk melihat integrasi data pada *access log* dengan penambahan fitur pada *error log* dalam memprediksi serangan.
3. Untuk melihat efisiensi kedua algoritma tersebut dan mencari algoritma yang terbaik dalam memprediksi serangan dan bukan serangan pada *log web server* UPNVJ.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini ialah.

1. Untuk mengetahui hasil prediksi sistem dalam memprediksi sebuah serangan yang ada pada *log web server*.
2. Untuk membandingkan akurasi, presisi dan *recall* dari algoritma *Hidden Markov Model* dan *Naïve Bayes* dalam prediksi sebuah serangan atau tidak serangan pada dataset *log web server* UPNVJ.

1.5 Batasan Masalah

Batasan masalah yang ada pada penelitian ini yaitu.

1. Mendeteksi serangan hanya berfokus pada *log* dari *access log* dan *error log* dari *web server* Apache pada *website* SIAKAD UPNVJ dengan jangka waktu September 2019.
2. Menggunakan Categorical Naïve Bayes dalam pengolahan *dataset.*
3. Pengelolaan data menggunakan bahasa pemrograman *Python*.
4. Penelitian hanya berfokus untuk perbandingan hasil akhir dari tiap-tiap algoritma yang digunakan.
5. Penelitian akan memiliki hasil akhir berupa algoritma mana yang paling baik serta efisien dalam melakukan klasifikasi serangan dan bukan serangan.

1.6 Luaran yang Diharapkan

Luaran yang diharapkan pada penelitian ini ialah hasil prediksi algoritma *Naïve Bayes* melalui *access log* dan *error log* apakah terdapat sebuah serangan atau tidak dan akurasi, presisi serta *recall* yang dihasilkan lebih baik atau tidak dari algoritma *Hidden Markov Model* terhadap *access log* dan *error log.*

1.7 Sistematika Penulisan

Berikut merupakan sistematika penulisan berupa gambaran secara terperinci mengenai tiap bab pada penulisan yang menjelaskan kesinambungan tiap bab satu sama lain yang akan dijelaskan sebagai berikut :

BAB 1 : PENDAHULUAN

Pada Bab ini berisi Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Batasan Masalah, Luaran yang Diharapkan, dan Sistematika Penulisan.

BAB 2 : LANDASAN TEORI

Pada Bab II Landasan Teori berisi tentang teori-teori mendasar, referensi jurnal, dan metode yang digunakan dalam penelitian ini.

BAB 3 : METODOLOGI PENELITIAN

Pada Bab III Metodologi Penelitian berisi tentang kerangka pikir, alur metode dalam memproses penelitian ini, serta segala metode yang terdapat dalam penelitian ini.

BAB 4 : HASIL DAN PEMBAHASAN

Pada Bab IV Hasil dan Pembahasan berisi tentang penjelasan mengenai proses pengolahan data dan pembuatan model untuk sistem, lalu pembahasan tentang analisis hasil pengujian dari data serta hasil perbandingan dengan penelitian terdahulu.

BAB 5 : PENUTUP

Pada Bab V Penutup berisi tentang kesimpulan dari hasil penelitian dan perbandingan yang dilakukan pada bab 4 (empat) serta saran yang dapat digunakan sebagai acuan untuk penelitian selanjutnya agar sistem dapat diperbaharui lebih baik dan lebih dinamis.

DAFTAR PUSTAKA