BAB 1
PENDAHULUAN

1.1 Latar Belakang

Aplikasi web sebagai media pertukaran informasi dengan konsep *client* dan *server*. Web memiliki peranan penting dalam perkembangan teknologi pada saat ini, hampir seluruh lapisan masyarakat sudah memanfaatkan web dalam kehidupan sehari-hari, seperti media *social, e-commerce,* kursus online, periklanan, internet banking dan sebagainya (Veerasamy, 2010)(Chaudhari & Vaidya, 2014)(Gupta & B. B. Gupta, 2017). Keadaan ini juga membuat banyak orang tertarik untuk mengambil keuntungan secara ilegal dengan memanfaatkan kelemahan-kelemahan yang ada pada teknologi web. Pelaku yang melakukan kegiatan tersebut disebut *hacker*. Menurut *Open Web Application Security Project* (*OWASP*), salah satu serangan yang paling sering dilakukan *hacker* untuk menyerang aplikasi web adalah *Cross-site Scripting* (XSS) (Pannu, 2014)(Anton, Manico, & Bird, 2016)(Wiliams & Wichers, 2017) dan menurut Bridgwater, sebanyak 68% aplikasi web dari seluruh dunia rentan terhadap serangan XSS (Bridgwater, 2016). XSS merupakan serangan yang memanfaatkan celah keamanan untuk memasukkan *malicious script* ke dalam halaman website, *script* tersebut selanjutnya akan mengarahkan *user* ke website yang telah dirancang untuk dapat mengambil *cookies* atau *session* yang dimiliki *user*. XSS umumnya dibagi menjadi 2, yaitu *reflected XSS* dan *stored XSS* (Almudena Alcaide Raya, Jorge Blasco Alis & Diaz-Pabón, 2011)(Gupta & B. B. Gupta, 2017). *Reflected XSS* adalah serangan XSS yang dilakukan dengan cara menyisipkan *script-script*

Javascript berbahaya kedalam url. Sedangkan *stored XSS* adalah serangan XSS yang dilakukan dengan cara menyisipkan *script-script Javascript* berbahaya ke dalam *database.*

Salah satu upaya pencegahan terhadap serangan XSS ini adalah dengan metode deteksi menggunakan *machine learning*. *Machine learning* adalah metode untuk menganalisa pola dari data-data yang ada berdasarkan parameter-parameter pembeda atau biasa disebut fitur. Metode ini menggunakan pola-pola yang telah didaftarkan untuk mengenali *script-script* berbahaya yang umumnya digunakan dalam serangan XSS.

Penelitian yang ada pada saat ini hanya mengecek keberadaan *tag script* pada url. Sehingga machine learning mengalami kesulitan ketika *tag* script disusupi oleh beberapa string. Oleh karena itu, penulis mengusulkan sebuah metode pendeteksi serangan XSS menggunakan kombinasi *machine learning* dan metode n-gram. Adapun metode *machine learning* yang digunakan adalah *SVM, KNN* dan *Naïve Bayes,* ciri-ciri dari metode ini adalah menemukan fungsi pemisah (klasifier) yang optimal yang bisa memisahkan dua set data dari dua kelas yang berbeda (Munawarah, Soesanto, & Faisal, 2016). Metode n-gram adalah metode untuk mendeteksi kemiripan antara 2 kalimat (Lisangan, 2013). Dengan metode n-gram, penulis akan mencari kemiripan antara url yang ada pada data *training* dengan *script* berbahaya. Penambahan metode n-gram diharapkan dapat memperkuat pendeteksian terhadap serangan XSS khusus nya pada fitur *tag script* yang semakin bervariasi namun memiliki sumber atau pembuat yang sama.

1.2 Rumusan Masalah

Secara umum, pertanyaan yang dapat dijawab setelah penelitian ini selesai dilakukan adalah sebagai berikut :

1. Seberapa efektif metode *machine learning* dan n-gram dapat mendeteksi serta mampu menentukan suatu url positif atau negatif megandung serangan XSS ?
2. Apakah metode n-gram bisa digunakan untuk mencari kemiripan url yang terinjeksi dengan XSS dengan data-data yang ada ?

1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah :

1. Untuk menentukan metode *machine learning* manakah yang terbaik ketika metode tersebut di gabungkan dengan mentode n-gram.
2. Untuk memperkuat pendeteksian serangan XSS dengan variasi serangan XSS terbaru yang ada pada saat ini menggabungkan metode *machine learning* dan n-gram sehingga kedepannya dapat dikembangkan untuk menjadi *extension browser* guna mencegah serangan XSS pada pengguna internet.

Manfaat dari penelitian ini adalah:

1. Mencegah pengguna internet terjebak dalam serangan XSS dalam menjelejahi internet melalui browser.
2. Menjadikan serangan XSS bukan suatu ancaman bagi pengguna internet.
3. Menjadi solusi terhadap variasi serangan XSS yang terjadi saat ini.

1.4 Ruang Lingkup Penelitian

1. Untuk pengujian, *data training* yang digunakan diambil dari http://xssed.com sebanyak 200 website.
2. Pada penelitian ini metode yang digunakan adalah metode *Machine Learning* SVM,KNN dan Naïve Bayes.
3. Menggunakan metode n-gram sebagai metode tambahan untuk memperkuat metode sebelumnya.