ABSTRAK

Twitter merupakan *platform* media sosial yang menjadi tempat bagi banyak orang untuk dapat mengunggah berbagai hal, tidak terkecuali unggahan yang mengandung unsur ancaman keamanan suatu sistem. Tentunya ini merupakan hal yang berbahaya jika seseorang mengunggah celah keamanan suatu sistem. Ancaman sistem yang dipublikasi dapat disalah gunakan oleh orang lain sehingga merugikan pemilik sistem.

Untuk mengantisipasi hal ini, maka dibuat sistem untuk mendeteksi unggahan yang mengandung unsur ancaman (*threat*) dan kerentanan (*vulnerability*) sistem pada media sosial Twitter. Sistem ini menerapkan algoritma *text processing* yang menggunakan metode Naïve Bayes dan TF-IDF (*Term Frequency – Inverse Document Frequency*). Metode ini dipilih karena dianggap dapat menghasilkan akurasi yang baik meskipun dengan data training yang sedikit.

Pada penelitian Tugas Akhir ini, hasil akhir yang didapatkan adalah sistem dapat membedakan *tweet* yang mengandung unsur *threat* atau *vulnerability*, dan yang tidak. Dengan rasio pembagian dataset ke dalam data training dan data testing adalah 70%:30% dan 80%:20%, keduanya mendapatkan nilai akurasi sebesar 88%, nilai presisi sebesar 88%, *recall* sebesar 88%, dan *F1 score* sebesar 88%.

Kata Kunci: *text mining*, Naïve Bayes, TF-IDF, *threat*, *vulnerabilities*, klasifikasi teks.