ABSTRACT

COMPARATIVE ANALYSIS OF DECISION TREE AND NAÏVE BAYES ALGORITHM IN MALWARE DETECTION WITH VARIABLE DISCRITIZATION

by

Agus Navirgo

navirgo.eldrich@gmail.com

By the more sophisticated infrastructure, devices and digital technology, people nowadays are living side by side with cyber or the internet. The internet not only has various functions and roles, but also has negative impact; namely cybercrime, one of which is malware. In Indonesia, the threat of cybercrime mostly targets the financial, insurance and property industry sectors, then manufacturing, constructions, and services. Therefore, malware requires special attention. Malware detection techniques are required to determine malware threats, one of which is Classification Algorithm. This study aims to analyze the comparison of the Decision Tree C4.5 Algorithm with the Naïve Bayes on malware detection with variable discretization. The results of the study found that the use of the Decision Tree C4.5 Algorithm is better than the Naïve Bayes. This is proven by the accuracy rate starting from 88.20% to 100%; while the Naïve Bayes starting from 69.60% to 92.10%. These results were obtained by using the Tools Rapidminer Studio with 10-Fold Cross Validation.

Keywords : Malware, Decision Tree, Naïve Bayes, Discretized

INTISARI

ANALISIS PERBANDINGAN ALGORITMA DECISION TREE DAN NAÏVE BAYES PADA PENDETEKSIAN MALWARE DENGAN DISKRITISASI VARIABEL

oleh

Agus Navirgo

navirgo.eldrich@gmail.com

Dengan semakin canggihnya infrastruktur, perangkat dan teknologi digital, kini semakin banyak masyarakat hidup berdampingan dengan dunia maya atau internet. Internet mempunyai fungsi dan peran yang beragam serta menimbulkan dampak negatif yaitu kejahatan siber salah satunya malware. Di Indonesia, ancaman kejahatan siber berupa malware paling banyak menyasar sektor industri keuangan, asuransi, dan properti, diikuti manufaktur, konstruksi, serta jasa. Oleh karena itu malware membutuhkan perhatian khusus, karena sebagian besar sumber ancaman di ruang siber berasal dari malware. Untuk mendeteksi ancaman malware tersebut diperlukan teknik pendeteksian malware salah satunya dengan algoritma klasifikasi data mining. Penelitian ini bertujuan untuk menganalisa perbandingan algoritma Decision Tree C4.5 dengan Naïve Bayes pada pendeteksian malware dengan diskritisasi variabel. Dari hasil penelitian ditemukan bahwa penggunaan algoritma decesion Tree C4.5 lebih baik dari naïve bayes, hal ini dibuktikan dengan tingkat akurasi dimulai 88,20 % - 100 %, sedangkan naïve bayes dimulai 69,60 % - 92,10 %. Hasil tersebut diperolah menggunakan Tools Rapidminer Studio dengan 10 fold cross validation.

Kata Kunci : Malware, Decision Tree, Naïve Bayes, Diskritisasi