ANALISIS KLASIFIKASI HOAX PADA *UNSTRUCTURED DATA TEXT* DI SITUS PORTAL BERITA DETIK.COM DAN TURNBACKHOAX.ID

Achmad Kurniansyah Thalib

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Islam Indonesia

INTISARI

Perkembangan arus informasi yang begitu cepat yang ditimbulkan oleh perkembangan teknologi informasi dapat menyebabkan informasi berupa berita beredar luas secara cepat. Penyebaran berita yang sangat cepat dapat menyebabkan penyaringan informasi sangat sulit dilakukan. Lajunya arus informasi memberikan dampak yang luas pada masyarakat, bahkan mampu masuk dalam hingka ke sisi psikologi-sosial dan berpotensi menjadi ancaman besar bagi keutuhan Bangsa. Informasi yang dengan mudah didapat sangat lemah dalam hal validitas dan tidak jarang merupakan sebuah hoax. Data yang digunakan merupakan judul berita dari situs detik.com dan turnbackhoax.id. Analisis untuk mengklasifikasikan hoax menggunakan Naive Bayes Classifier dan Support Vector Machine. Hasil yang diperoleh adalah Untuk metode Naive Bayes Classifier dengan data training sebesar 1.480 dan data test sebanyak 369 diperoleh akurasi sebesar 85,09% dan untuk metode Support Vector Machine diperoleh akurasi 83,74%. Selain itu, Dari hasil menggali informasi dengan text mining, didapatkan kata yang paling banyak untuk kategori berita adalah "Harga", diikuti dengan "KPK", "Saham", "Indonesia", "DPR", dan "Polisi". Untuk kategori hoax, kata-kata yang paling banyak adalah kata "Harga", diikuti dengan "KPK", "Saham", "Indonesia", "DPR", dan "Polisi".

Kata Kunci : *Berita, Hoax, Text Mining, SVM, NBC.*

HOAX CLASSIFICATION ANALYSIS ON UNSTRUCTURED DATA TEXT ON PORTAL SITES DETIK.COM NEWS AND TURNBACKHOAX.ID

Achmad Kurniansyah Thalib

Department Statistics, Faculty of Mathematics and Natural Science

Islamic University of Indonesia

ABSTRACT

The rapid flow of information generated by the propagation of information technology can lead to information in the form of gloomy news quickly. Spreading the news very easily can lead to very difficult information filtering. The flow of information that provides broad benefits to society, can even enter into the psychology and social for the integrity of the Nation. Information that is easily obtained is extremely dangerous in terms of validity and is not uncommonly a hoax. Data which is the title of news from detik.com site and turnbackhoax.id. To classify hoaxes using the Naive Bayes Classifier and Support Vector Machine. The results obtained are for NBC method with training data of 1.480 and test data of 369 obtained by 85.09% and for SVM method obtained an accuracy of 83.74%. In addition, From the merging of information with text mining, the keyword for the news category is "Price", followed by "KPK", "Stock", "Indonesia", "DPR", and "Police". For the hoax category, the most words are the word "Price", followed by "KPK", "Stock", "Indonesia", "DPR", and "Police" .

Key Word : *News, Hoax, NBC, Text Mining , SVM*