

## ABSTRAK

**Agus Parwata** (2019), Analisis Tingkat Publikasi Data Pribadi Sensitif Masyarakat pada Situs Resmi Komisi Pemilihan Umum dengan Metode TF-IDF dan Metode Naïve Bayes. Tesis: Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Tesis ini telah disetujui dan diperiksa oleh Pembimbing I: Kadek Yota Ernanda Aryanto, Ph.D. dan Pembimbing II: Dr. Dewa Gede Hendra Divayana.

**Kata-kata kunci:** Data Pribadi Sensitif, *Web Crawling*, *Data Mining*, TF-IDF, Naïve Bayes

Pelindungan data pribadi sensitif masyarakat di Indonesia merupakan hal yang sangat penting karena banyak ditemukan data pribadi masyarakat yang dipublikasi oleh instansi pemerintahan tanpa mempertimbangkan aturan yang ada. Pada penelitian ini dilakukan uji pencarian data berupa *hyperlink* dan kontennya pada situs resmi Komisi Pemilihan Umum (KPU) Pusat dan Provinsi menggunakan bantuan *Web Crawling*. Konten yang telah diperoleh dilakukan praproses data dengan bantuan *Text Preprocessing* dan kemudian dilakukan pembobotan dengan metode TF-IDF untuk selanjutnya diklasifikasikan menggunakan metode Naïve Bayes. Setelah itu dilakukan analisis terkait jenis data pribadi sensitif yang dipublikasikan serta tingkat publikasi berdasarkan kelompok wilayah (pusat, timur, tengah, barat), serta penyusunan rekomendasi yang dapat dilakukan oleh KPU dalam melakukan publikasi data pribadi berdasarkan UU Nomor 23 Tahun 2006 dan Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016. Berdasarkan 6.700 *instance* yang dianalisis ditemukan sebanyak 6.430 data pribadi yang dipublikasi. Jenis data pribadi yang dipublikasikan adalah nama lengkap, tempat lahir, tanggal lahir, jenis kelamin, agama, status perkawinan, NIP (nomor induk pegawai), NIK (nomor induk kependudukan), NPWP (nomor pokok wajib pajak), nomor rekening, nomor telepon/ *handphone*, alamat *email*, alamat tempat tinggal, jabatan/pekerjaan, foto wajah. Tingkat publikasi berdasarkan jumlah data yang ditemukan yaitu: KPU Pusat sebesar 11.45 %, wilayah timur sebesar 21.60 %, wilayah tengah sebesar 17.01 %, dan wilayah barat sebesar 49.94 %. Akurasi dari metode Naïve Bayes rata-rata sebesar 96.99 %. Rekomendasi yang harus dilakukan KPU dalam melakukan publikasi adalah menghormati data pribadi seseorang sebagai sebuah privasi, setiap publikasi data harus berdasarkan persetujuan dan menyediakan narahubung (*contact person*) yang mudah dihubungi.

## ABSTRACT

**Agus Parwata (2019)**, *Publication Level Analysis of Public Sensitive Personal Data on the Official Site of the General Election Commission Using TF-IDF method and Naïve Bayes method. Thesis, Computer Science, Post Graduate Study Program, Ganesha University of Education.*

*This thesis has been supervised and approved by Supervisor I: Kadek Yota Ernanda Aryanto, Ph.D. and Supervisor II: Dr. Dewa Gede Hendra Divayana.*

**Keywords:** *Sensitive Personal Data, Web Crawling, Data Mining, TF-IDF, Naïve Bayes*

*The protection of sensitive personal data in Indonesia is very important because many public personal data are found published by government agencies without considering existing rules. In this study a data search test was conducted in the form of hyperlinks and their content on the official website of the Central and Provincial General Elections Commission (KPU) using Web Crawling assistance. The content that has been obtained is preprocessed with the help of Text Preprocessing and then weighted using the TF-IDF method to be further classified using the Naïve Bayes method. After that, an analysis is made regarding the type of sensitive personal data published and the level of publication based on regional groups (central, east, central, west), as well as the preparation of recommendations that can be carried out by the KPU in publishing personal data based on Law Number 23 Year 2006 and the Minister of Communications Regulation and Informatics Number 20 Year 2016. Of 6,700 instances analyzed, 6,430 personal data were published. Types of personal data published are full name, place of birth, date of birth, gender, religion, marital status, nip (employee identification number), nik (residence registration number), npwp (taxpayer identification number), account number, telephone number / mobile, email address, residential address, position / job, face photo. The level of publication is based on the amount of data found, namely: the Central KPU is 11.45%, the eastern region is 21.60%, the middle region is 17.01%, and the western region is 49.94%. The accuracy of the Naïve Bayes method averages 96.99%. The recommendation that must be made by KPU in making publications is to respect a person's personal data as a privacy, each data publication must be based on agreement and provide a contact person who is easily contacted.*