

ABSTRAK

Alit Sudrastawa, I Putu (2019), Analisis Tingkat Publikasi Data Pribadi Sensitif pada Situs *Web* Sistem Informasi LPTK di Indonesia dengan Metode *Expectation Maximization*. Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Tesis ini telah disetujui dan diperiksa oleh Pembimbing I: Kadek Yota Ernanda Aryanto, S.Kom., M.T., Ph.D. dan Pembimbing II: Prof. Drs. Sariyasa, M.Sc., Ph.D.

Kata-kata kunci: Data Pribadi Sensitif, *Personally Identifiable Information*, *Data Mining*, *Clustering*, *Expectation Maximization*

Data pribadi adalah segala jenis data yang merujuk pada suatu individu. Data tersebut perlu disimpan, dirawat, dan dijaga kebenarannya serta dilindungi kerahasiaannya guna mencegah eksploitasi dan penyalahgunaan oleh pihak yang tidak bertanggung jawab. Pada penelitian ini dilakukan analisis terhadap jenis data pribadi sensitif yang dipublikasikan beserta sebarannya pada situs *web* sistem informasi LPTK di Indonesia, selanjutnya dilakukan analisis *clustering* dengan metode *Expectation Maximization*. Hasilnya dari 72.522 *instance* yang dianalisis berdasarkan kriteria dari NIST dan Undang-Undang Nomor 23 Tahun 2006 ditemukan sebanyak 189.358 data pribadi sensitif yang 87,72% diantaranya merupakan *Critical personally identifiable information* (PII) dan sisanya 12,28% merupakan *Potential PII*. Jenis *Critical PII* yang dipublikasikan mencakup tempat lahir, tanggal lahir, alamat rumah, nomor telepon/faks, alamat surel, foto wajah, agama, dan NIP, sedangkan *Potential PII* mencakup jabatan, lokasi/unit kerja, wilayah kabupaten tempat tinggal, dan usia. Akurasi *clustering* rata-rata yang diperoleh dengan metode *Expectation Maximization* adalah 98,53% dengan kesalahan hasil klaster sebanyak 1.682 *instance* atau sekitar 2,32%. Terkait pengembangan penelitian di masa mendatang, kriteria penentuan data pribadi sensitif yang digunakan dapat merujuk pada regulasi lokal sehingga hasil deskripsi yang diperoleh dan rekomendasi yang disusun dapat lebih sesuai untuk wilayah Indonesia.

ABSTRACT

I Putu Alit Sudrastawa (2019), *Publication Level Analysis of Sensitive Personal Data in LPTK's Information System Website in Indonesia Using Expectation Maximization 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, S.Kom., M.T., Ph.D. and Supervisor II: Prof. Drs. Sariyasa, M.Sc., Ph.D.

Keywords: *Sensitive Personal Data, Personally Identifiable Information, Data Mining, Clustering, Expectation Maximization*

Personal data is described as any information that relates to an identifiable individual which is stored, maintained its truth, and is protected its confidentiality to prevent abuse by irresponsible parties. In this study, we analyzed the types of sensitive personal data published along with their distribution on the LPTK's information system websites. We also implemented data mining clustering analysis using the Expectation Maximization method. As a result, from 72,522 instances that have been analyzed based on NIST criteria and Indonesian Regulation Number 23 of the year 2006, around 87.72% out of 189,358 obtained sensitive personal data were Critical personally identifiable information (PII), while the remaining 12.28% were Potential PII. Types of Critical PII published including the place of birth, date of birth, home address, telephone number, email address, face photo, religion, and employee identification number. Meanwhile, the Potential PII including position, work location/unit, district area of residence, and age. Average clusters accuracy obtained using Expectation Maximization was 98.53% with 1,682 incorrectly clustered instances. In the future, the used criteria for determining sensitive personal data could refer to local regulations so that the results and the recommendations obtained are more suitable to the territory of Indonesia.