

**ANALISIS TINGKAT PUBLIKASI DATA PRIBADI
SENSITIF PADA SITUS *WEB* SISTEM INFORMASI
LPTK DI INDONESIA DENGAN METODE
*EXPECTATION MAXIMIZATION***

TESIS

**Diajukan kepada
Universitas Pendidikan Ganesha
untuk Memenuhi Sebagian Persyaratan
Memperoleh Gelar Magister Ilmu Komputer
Program Studi Ilmu Komputer**

oleh

**I Putu Alit Sudrastawa
NIM 1729101026**



**PROGRAM STUDI ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS PENDIDIKAN GANESHA**

Juli 2019

Tesis oleh I Putu Alit Sudrastawa ini telah diperiksa dan disetujui untuk mengikuti Ujian Tesis.

Singaraja, 15 Juli 2019

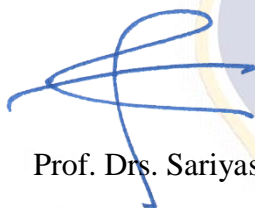
Pembimbing I,



Kadek Yota Ernanda Aryanto, S.Kom., M.T., Ph.D.

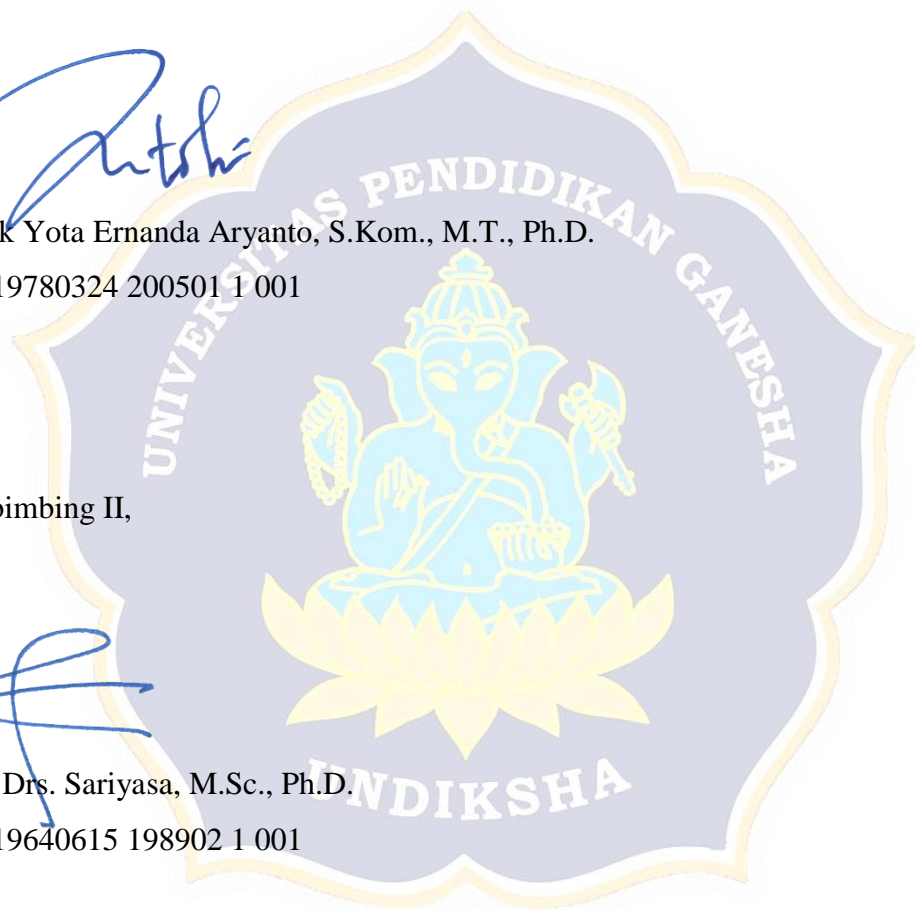
NIP 19780324 200501 1 001

Pembimbing II,



Prof. Drs. Sariyasa, M.Sc., Ph.D.


NIP 19640615 198902 1 001





Tesis oleh I Putu Alit Sudrastawa ini telah berhasil dipertahankan di depan tim penguji dan dinyatakan diterima sebagai sebagian persyaratan untuk memperoleh gelar Magister Ilmu Komputer di Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.


Disetujui pada tanggal : 1 Agustus 2019


Oleh Tim Penguji


....., Ketua (Dr. Gede Indrawan, S.T., M.T.)
NIP 19760102 200312 1 001


....., Anggota (Dr. I Gede Aris Gunadi, S.Si., M.Kom.)
NIP 19770318 200812 1 004


....., Anggota (Dr. Dewa Gede Hendra Divayana, S.Kom., M.Kom.)
NIP 19840724 201504 1 002



....., Anggota (Prof. Drs. Sariyasa, M.Sc., Ph.D.)
NIP 19640615 198902 1 001


....., Anggota (Kadek Yota Ernanda A., S.Kom, M.T., Ph.D.)
NIP 19780324 200501 1 001

Mengetahui Direktur

Program Pascasarjana Undiksha,




Prof. Dr. I Gusti Putu Suharta, M.Si.

NIP 19621215 198803 1 002

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Ilmu Komputer dari Program Pascasarjana Universitas Pendidikan Ganesha seluruhnya merupakan hasil karya saya sendiri. Bagian-bagian tertentu dalam penulisan tesis yang saya kutip dari karya orang lain telah ditulis sumbernya secara jelas dan sesuai dengan norma, kaidah, serta etika akademis.

Apabila di kemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku di wilayah Negara Kesatuan Republik Indonesia.

Singaraja, 15 Juli 2019

Yang memberi pernyataan,



I Putu Alit Sudrastawa

PRAKATA

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas rahmat-Nya penulis dapat menyelesaikan tesis yang berjudul “Analisis Tingkat Publikasi Data Pribadi Sensitif pada Situs *Web* Sistem Informasi LPTK di Indonesia dengan Metode *Expectation Maximization*” tepat pada waktunya. terselesainya tesis ini tidak lepas dari bantuan banyak pihak yang telah memberikan masukan serta dukungan kepada penulis. Untuk itu, penulis mengucapkan terima kasih kepada:

- 1) Bapak Dr. Gede Indrawan, S.T., M.T. selaku Kaprodi Ilmu Komputer yang telah memotivasi penulis dalam menyelesaikan pendidikan S2 di Program Studi Ilmu Komputer Universitas Pendidikan Ganesha;
- 2) Bapak Kadek Yota Ernanda A., S.Kom, M.T., Ph.D. selaku Pembimbing I dan Bapak Prof. Drs. Sariyasa, M.Sc., Ph.D. selaku Pembimbing II yang telah memberikan petunjuk selama penyusunan tesis ini;
- 3) Bapak Dr. I Gede Aris Gunadi, S.Si., M.Kom. selaku Penguji I dan Bapak Dr. Dewa Gede Hendra Divayana, S.Kom., M.Kom. selaku Penguji II yang juga telah memberikan arahan dalam penyelesaian tesis ini;
- 4) Ni Komang Candra Wahyuni, pasangan penulis yang tiada hentinya memberikan dukungan dan semangat kepada penulis;
- 5) Serta keluarga penulis dan rekan-rekan yang telah memberikan bantuan berupa masukan dan dukungan bagi penulis.

Penulis menyadari bahwa laporan tesis ini masih jauh dari sempurna. Oleh sebab itu, segala kritik dan saran sangat penulis hargai. Akhir kata, penulis berharap semoga laporan tesis ini bermanfaat bagi pembaca dan dunia pendidikan.

Singaraja, 15 Juli 2019

Penulis,

I Putu Alit Sudrastawa

DAFTAR ISI

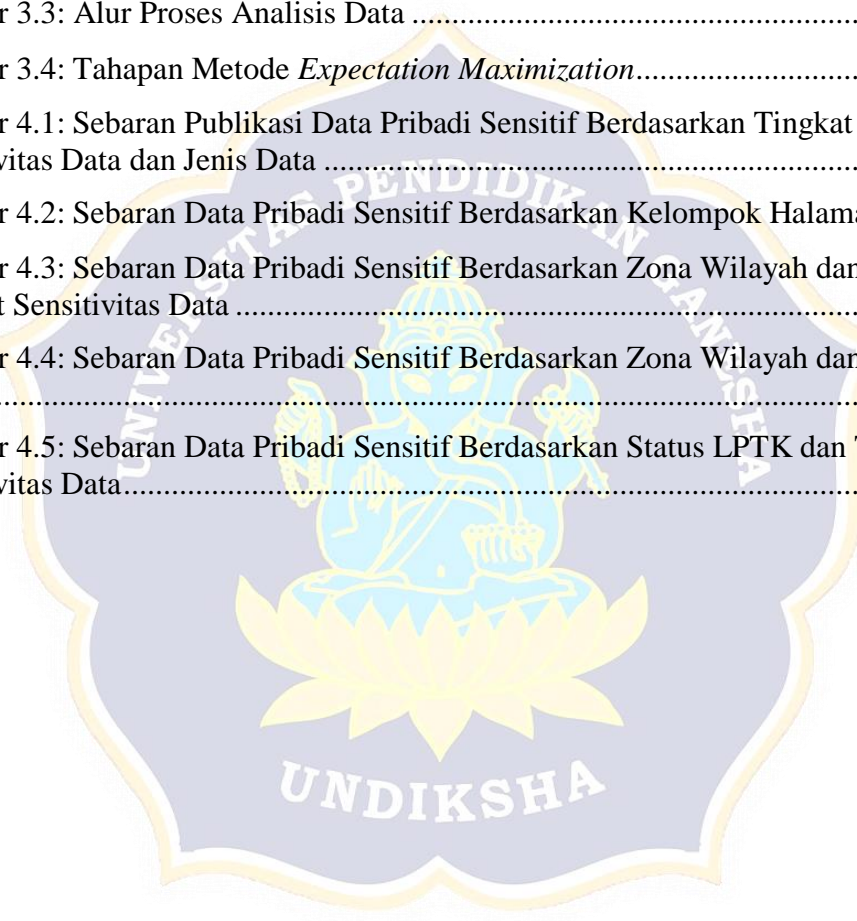
	(halaman)
LEMBAR PERNYATAAN	i
PRAKATA	ii
ABSTRAK	iii
<i>ABSTRACT</i>	iv
DAFTAR ISI	v
DAFTAR GAMBAR	vii
DAFTAR TABEL	viii
BAB I - PENDAHULUAN	1
1.1. Latar Belakang Penelitian	1
1.2. Identifikasi Masalah	5
1.3. Batasan Masalah	6
1.4. Rumusan Masalah	7
1.5. Tujuan Penelitian	8
1.6. Manfaat Penelitian	8
BAB II - KAJIAN PUSTAKA, PENELITIAN SEBELUMNYA, DAN KERANGKA KONSEP	10
2.1. Kajian Pustaka	10
1) LPTK di Indonesia	10
2) Sistem Elektronik dan Sistem Informasi	11
3) Data Pribadi Sensitif	12
4) Privasi	14
5) Perlindungan Data Pribadi	16
6) <i>Web Crawling</i>	16
7) Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 ..	18
8) <i>Data Mining</i>	20
9) <i>Cross Validation</i>	27
2.2. Penelitian Sebelumnya	28
2.3. Kerangka Konsep	33
BAB III - METODE PENELITIAN	39

3.1.	Rancangan Penelitian.....	39
3.2.	Subjek Penelitian.....	40
3.3.	Metode Pengumpulan Data.....	41
3.4.	Metode <i>Preprocessing</i> Data.....	43
3.5.	Metode Analisis Data.....	47
3.6.	Metode <i>Expectation Maximization</i>	50
BAB IV - HASIL DAN PEMBAHASAN		60
4.1.	Deskripsi Data.....	60
4.2.	Hasil Analisis	60
4.3.	Pembahasan.....	80
4.4.	Rekomendasi Berdasarkan Permenkominfo Nomor 20 Tahun 2016.....	82
4.5.	Rekomendasi Teknis	84
BAB V - PENUTUP		87
5.1.	Rangkuman	87
5.2.	Simpulan	88
5.3.	Saran.....	90
DAFTAR PUSTAKA		91



DAFTAR GAMBAR

	(halaman)
Gambar 2.1: Alur Proses <i>Web Crawling</i>	18
Gambar 2.2: Proses Ekstraksi Pola	22
Gambar 3.1: Alur Proses Pengumpulan Data	43
Gambar 3.2: Tahapan <i>Preprocessing</i> Data	46
Gambar 3.3: Alur Proses Analisis Data	49
Gambar 3.4: Tahapan Metode <i>Expectation Maximization</i>	51
Gambar 4.1: Sebaran Publikasi Data Pribadi Sensitif Berdasarkan Tingkat Sensitivitas Data dan Jenis Data	62
Gambar 4.2: Sebaran Data Pribadi Sensitif Berdasarkan Kelompok Halaman	65
Gambar 4.3: Sebaran Data Pribadi Sensitif Berdasarkan Zona Wilayah dan Tingkat Sensitivitas Data	66
Gambar 4.4: Sebaran Data Pribadi Sensitif Berdasarkan Zona Wilayah dan Jenis Data	66
Gambar 4.5: Sebaran Data Pribadi Sensitif Berdasarkan Status LPTK dan Tingkat Sensitivitas Data.....	68



DAFTAR TABEL

	(halaman)
Tabel 2.1. Hasil Uji Awal Terhadap Sampel Penelitian Pendahuluan.....	33
Tabel 3.1. Fitur Data Masukan.....	44
Tabel 3.2. <i>Keyword</i> Pemberian Nilai	45
Tabel 3.3. Kelas Klaster	49
Tabel 3.4. Kondisi Awal Probabilitas Klaster C Terhadap Fitur X	52
Tabel 3.5. Frekuensi Fitur X	52
Tabel 3.6. Model Parameter	53
Tabel 3.7. Frekuensi Fitur X yang Baru	53
Tabel 3.8. Nilai Probabilitas Klaster C yang Baru	55
Tabel 3.9. Probabilitas Klaster C Terhadap Masing-Masing <i>Instance</i> pada Fitur X untuk Iterasi ke-1.....	57
Tabel 3.10. Nilai Frekuensi untuk Iterasi ke-1.....	57
Tabel 3.11. Model Parameter Baru untuk Iterasi ke-1.....	58
Tabel 3.12. Hasil Iterasi ke-19 Perhitungan <i>Expectation Maximization</i>	59
Tabel 4.1. Jenis Data Pribadi Sensitif yang Dipublikasikan	61
Tabel 4.2. Jumlah <i>Critical PII</i> dan <i>Potential PII</i> (Bernilai “ <i>true</i> ”) dari Masing Masing Kriteria Data Pribadi Sensitif	63
Tabel 4.3. Jumlah NonPII (Bernilai “ <i>false</i> ”) dari Masing-Masing Kriteria Data Pribadi Sensitif.....	64
Tabel 4.4. Sebaran Nilai “ <i>true</i> ” dari Masing-Masing Kriteria Berdasarkan Zona Wilayah	67
Tabel 4.5. Sebaran Nilai “ <i>true</i> ” Sensitif dari Masing-Masing Kriteria Berdasarkan Status LPTK.....	69
Tabel 4.6. Hasil <i>Clustering</i> dengan Metode <i>Expectation Maximization</i>	70
Tabel 4.7. Klaster yang Dihasilkan dengan Metode <i>Expectation Maximization</i> ..	71
Tabel 4.8. Perbandingan Nilai “ <i>true</i> ” dan “ <i>false</i> ” pada Masing-Masing Klaster..	73
Tabel 4.9. <i>Dataset</i> Universitas Y11	78
Tabel 4.10. <i>Dataset</i> Universitas Y1	79
Tabel 4.11. Golongan Data Pribadi Sensitif yang Dipublikasikan Berdasarkan Kriteria dari NIST	81