

**ANALISIS TINGKAT PUBLIKASI DATA PRIBADI
SENSITIF MASYARAKAT PADA SITUS RESMI
KOMISI PEMILIHAN UMUM DENGAN METODE TF-
IDF DAN METODE NAÏVE BAYES**

TESIS

oleh

Agus Parwata

NIM. 1729101027



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

2019

**ANALISIS TINGKAT PUBLIKASI DATA PRIBADI
SENSITIF MASYARAKAT PADA SITUS RESMI
KOMISI PEMILIHAN UMUM DENGAN METODE TF-
IDF DAN METODE NAÏVE BAYES**

TESIS

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

**oleh
Agus Parwata
NIM. 1729101027**



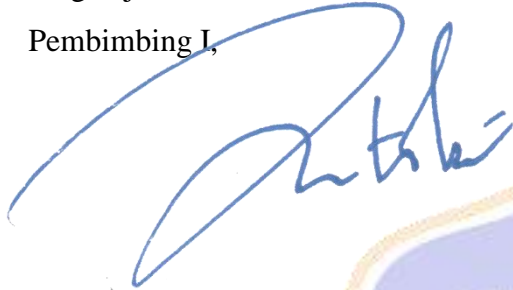
**PROGRAM STUDI ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS PENDIDIKAN GANESHA**

2019

Tesis oleh Agus Parwata ini telah diperiksa dan disetujui untuk mengikuti Ujian Tesis.

Singaraja, 24 Oktober 2019

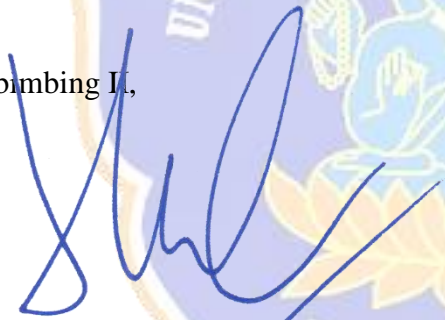
Pembimbing I,



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

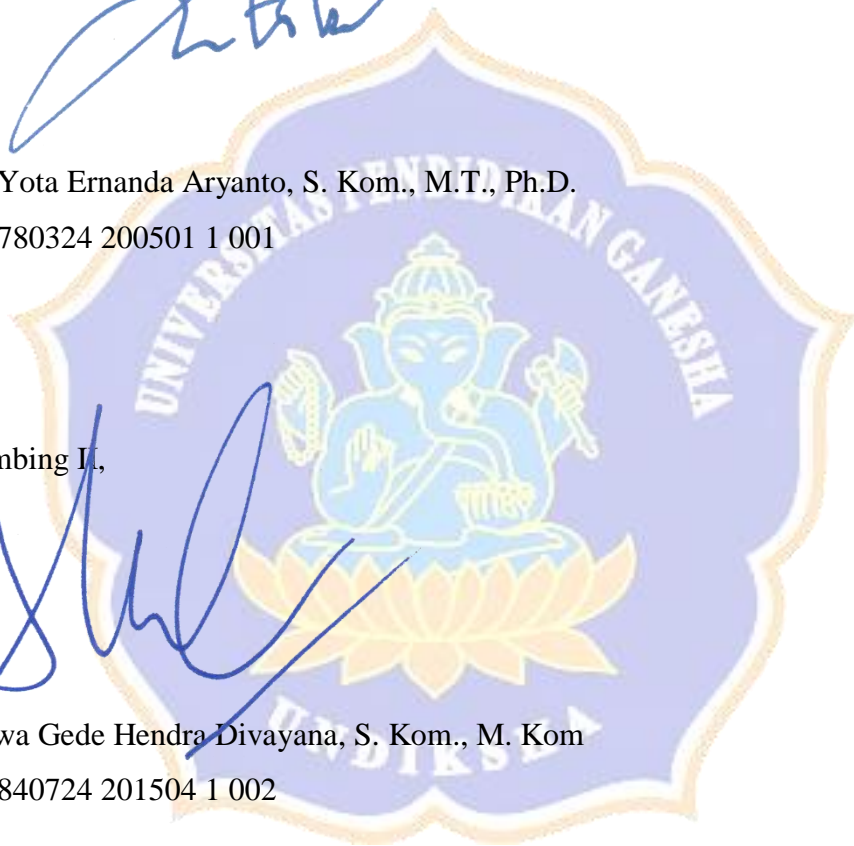
NIP 19780324 200501 1 001

Pembimbing II,



Dr. Dewa Gede Hendra Divayana, S. Kom., M. Kom


NIP 19840724 201504 1 002



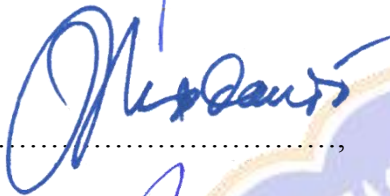
Tesis oleh Agus Parwata ini telah berhasil dipertahankan di depan tim penguji dan dinyatakan diterima sebagai sebagian persyaratan untuk memperoleh gelar Magister Ilmu Komputer di Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Disetujui pada tanggal : 24 Oktober 2019

Oleh Tim Penguji

.....



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

.....



Anggota (Prof. Dr. I Made Candiasa, MI. Komp.)
NIP 19601231 198601 1 004

.....


Anggota (Dr. Drs. I Nyoman Sukajaya, M.T)
NIP 19671115 199303 1 001

.....


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


.....


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

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.

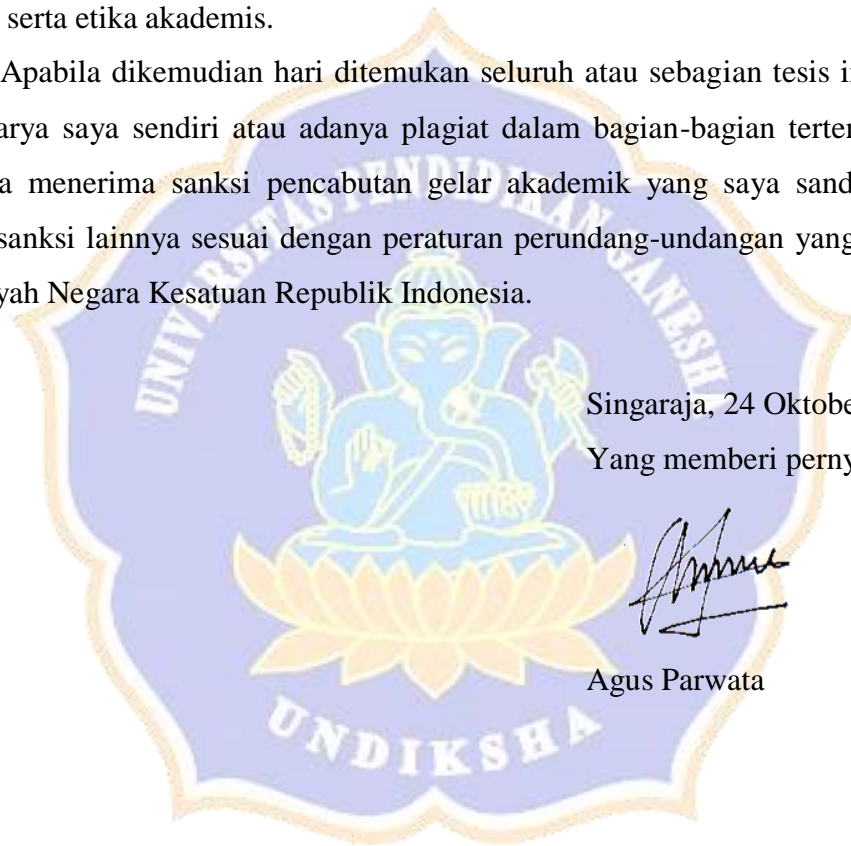
Apabila dikemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku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

Singaraja, 24 Oktober 2019

Yang memberi pernyataan,



Agus Parwata



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 Masyarakat pada Situs Resmi Komisi Pemilihan Umum dengan Metode TF-IDF dan Metode Naïve Bayes” tepat pada waktunya. terselesaikannya 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 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 Aryanto, Ph.D. selaku Pembimbing I dan Bapak Dr. Dewa Gede Hendra Divayana selaku Pembimbing II yang telah memberikan petunjuk selama penyusunan tesis ini;
- 3) Bapak Prof. Dr. I Made Candiasa selaku Penguji I dan Bapak Dr. I Nyoman Sukajaya selaku Penguji II yang juga telah memberikan arahan dalam penyelesaian tesis ini;
- 4) 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, 24 Oktober 2019



Agus Parwata

DAFTAR ISI

LEMBAR PERNYATAAN	i
PRAKATA	ii
ABSTRAK	iii
ABSTRACT	iv
DAFTAR ISI	v
DAFTAR GAMBAR	vii
DAFTAR TABEL	viii
DAFTAR LAMPIRAN	ix
BAB I	1
PENDAHULUAN	1
1.1. Latar Belakang Penelitian	1
1.2. Identifikasi Masalah	6
1.3. Batasan Masalah	7
1.4. Rumusan Masalah	8
1.5. Tujuan Penelitian	9
1.6. Manfaat Penelitian	10
BAB II	10
KAJIAN PUSTAKA, PENELITIAN YANG RELEVAN DAN KERANGKA KONSEP	10
2.1. Kajian Pustaka	10
1) Data Pribadi	10
2) Peraturan yang Terkait Data Pribadi	12
3) <i>Web Crawling</i>	14
4) <i>Data Mining</i>	16
5) Naïve Bayes	18
6) <i>Text Mining</i>	19
7) <i>Text Preprocessing</i>	19
2.2. Penelitian yang Relevan	21
2.3. Kerangka Konsep	26

BAB III.....	29
METODE PENELITIAN.....	29
3.1. Rancangan Penelitian.....	29
3.2. Data Penelitian.....	32
3.3. Pengumpulan Data.....	33
3.4. Praproses Data.....	34
3.5. Analisis data.....	48
3.6. Naïve Bayes.....	51
BAB IV.....	57
HASIL DAN PEMBAHASAN.....	57
4.1 Komisi Pemilihan Umum (KPU).....	57
4.2 Deskripsi Data.....	58
4.3 Hasil Analisis.....	59
4.4 Pembahasan.....	64
4.5 Rekomendasi Kepada KPU Terkait Publikasi Data Pribadi.....	76
4.6 Rekomendasi Teknis.....	78
BAB V.....	81
PENUTUP.....	81
5.1 Rangkuman.....	81
5.2 Simpulan.....	82
5.3 Saran.....	84
DAFTAR PUSTAKA.....	86

DAFTAR GAMBAR

Gambar 2.1 Cara Kerja <i>Web Crawler</i>	16
Gambar 3.1 Alur Proses Penelitian	31
Gambar 4.1. Jenis Data Pribadi yang Dipublikasikan Berdasarkan Pemilik Data	66
Gambar 4.2. Grafik Tingkat Publikasi Masing-Masing Provinsi Wilayah Timur	69
Gambar 4.3 Grafik Tingkat Publikasi Masing-Masing Provinsi Wilayah Tengah	70
Gambar 4.4. Grafik Tingkat Publikasi Masing-Masing Provinsi Wilayah Barat .	71
Gambar 4.5. Tingkat Publikasi Data Setiap Provinsi Pada Wilayah Timur Berdasarkan Jumlah Data yang Dipublikasikan	72
Gambar 4.6. Tingkat Publikasi Data Setiap Provinsi Pada Wilayah Tengah Berdasarkan Jumlah Data yang Dipublikasikan	73
Gambar 4.7. Tingkat Publikasi Data Setiap Provinsi Pada Wilayah Barat Berdasarkan Jumlah Data yang Dipublikasikan	74



DAFTAR TABEL

Tabel 3.1. Proses <i>Case Folding</i>	36
Tabel 3.2. Proses <i>Tokenization</i>	37
Tabel 3.3. Proses <i>Filtering</i>	38
Tabel 3.4. Contoh Dokumen yang akan Dibobotkan dengan TF-IDF.....	40
Tabel 3.5. Perhitungan Nilai TF dan DF.....	41
Tabel 3.6. Hasil Perhitungan Nilai IDF.....	43
Tabel 3.7. Hasil Perhitungan Nilai W.....	46
Tabel 3.8. <i>Data Training</i>	51
Tabel 3.9. Nilai Probabilitas Detail Variabel Dalam Kelas.....	53
Tabel 3.10. <i>Data Testing</i>	55
Tabel 3.11. Hasil Perkalian Semua Variabel Kelas Berdasarkan <i>Data Testing</i>	55
Tabel 3.12 Kelas Prediksi.....	56
Tabel 3.13. <i>Confusion Table</i>	56
Tabel 4.1 Jenis Data Pribadi Sensitif yang Dipublikasikan.....	60
Tabel 4.2 Tingkat Publikasi Berdasarkan Keseluruhan <i>Instance</i>	61
Tabel 4.3. Tingkat Publikasi Berdasarkan Data yang Ditemukan.....	61
Tabel 4.4. Akurasi Metode Naïve Bayes.....	62
Tabel 4.5. Tingkat Publikasi Masing-Masing Provinsi Wilayah Timur.....	68
Tabel 4.6. Tingkat Publikasi Masing-Masing Provinsi Wilayah Tengah.....	69
Tabel 4.7. Tingkat Publikasi Masing-Masing Provinsi Wilayah Barat.....	70

DAFTAR LAMPIRAN

Lampiran 1 Hasil Pengumpulan Data	90
Lampiran 2 Hasil Praproses Data dan Klasifikasi Manual	93
Lampiran 3 Hasil Pembobotan Data	95
Lampiran 4 Jenis Data Pribadi KPU Pusat dan Kelompok Wilayah Timur	96
Lampiran 5 Jenis Data Pribadi Kelompok Wilayah Tengah.....	97
Lampiran 6 Jenis Data Pribadi Kelompok Wilayah Barat.....	98

