

**KLASIFIKASI TINGKAT KEMACETAN LALU LINTAS BERBASIS  
ANALISIS *TWEET* MENGGUNAKAN *LATENT DIRICHLET  
ALLOCATION* DAN *N-GRAM***



**SKRIPSI**

**Disusun Sebagai Salah Satu Syarat  
untuk Memperoleh Gelar Sarjana Komputer  
pada Departemen Ilmu Komputer/ Informatika**

**Disusun Oleh :**

**MUHAMMAD IHSAN AJI WIEDJAYANTO**

**24010312140050**

**DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA**

**FAKULTAS SAINS DAN MATEMATIKA**

**UNIVERSITAS DIPONEGORO**

**2016**

**KLASIFIKASI TINGKAT KEMACETAN LALU LINTAS BERBASIS  
ANALISIS *TWEET* MENGGUNAKAN *LATENT DIRICHLET  
ALLOCATION* DAN *N-GRAM***



**SKRIPSI**

**Disusun Sebagai Salah Satu Syarat  
untuk Memperoleh Gelar Sarjana Komputer  
pada Departemen Ilmu Komputer/ Informatika**

**Disusun Oleh :**

**MUHAMMAD IHSAN AJI WIEDJAYANTO**

**24010312140050**

**DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA**

**FAKULTAS SAINS DAN MATEMATIKA**

**UNIVERSITAS DIPONEGORO**

**2016**

---

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Muhammad Ihsan Aji Wiedjayanto

NIM : 24010312140050

Judul : Klasifikasi Tingkat Kemacetan Lalu Lintas Berbasis Analisis *Tweet*  
Menggunakan *Latent Dirichlet Allocation* dan *N-Gram*.

Dengan ini saya menyatakan bahwa dalam tugas akhir/ skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, 16 Desember 2016



Muhammad Ihsan Aji Wiedjayanto

24010312140050

---

## HALAMAN PENGESAHAN

Judul : Klasifikasi Tingkat Kemacetan Lalu Lintas Berbasis Analisis *Tweet* Menggunakan *Latent Dirichlet Allocation* dan *N-Gram*.

Nama : Muhammad Ihsan Aji Wiedjayanto

NIM : 24010312140050

Telah diujikan pada sidang tugas akhir pada tanggal 2 Desember 2016 dan dinyatakan lulus pada tanggal 2 Desember 2016.

Semarang, 16 Desember 2016

Mengetahui,

Ketua Departemen Ilmu Komputer/ Informatika



Panitia Penguji Tugas Akhir

Ketua,

Helmie Arif Wibawa, S.Si./M.Cs

NIP. 197805162003121001

---

## HALAMAN PENGESAHAN

Judul : Klasifikasi Tingkat Kemacetan Lalu Lintas Berbasis Analisis *Tweet* Menggunakan  
*Latent Dirichlet Allocation* dan *N-Gram*.

Nama : Muhammad Ihsan Aji Wiedjayanto

NIM : 24010312140050

Telah diujikan pada sidang tugas akhir pada tanggal 2 Desember 2016.

Semarang, 16 Desember 2016

Pembimbing



Dr. Retno Kusumaningrum, S.Si, M.Kom

NIP. 198104202005012001

## ABSTRAK

Kemacetan lalu lintas merupakan permasalahan yang sering terjadi di kota-kota besar dan memberi dampak negatif terhadap lingkungan disekitarnya. *Twitter* sebagai salah satu media sosial, memungkinkan penggunaanya untuk berbagi informasi mengenai kemacetan lalu lintas disekitar mereka. Melihat permasalahan tersebut, pemanfaatan metode *Latent Dirichlet Allocation* (LDA) dapat diterapkan dalam klasifikasi *tweet*. Klasifikasi tingkat kemacetan lalu lintas berbasis analisis *tweet* dari pengguna *Twitter* bertujuan untuk menginformasikan kepada orang lain agar mengetahui lokasi-lokasi kemacetan. Saat identifikasi *tweet*, *bigram* akan diterapkan sebagai antisipasi penurunan kinerja klasifikasi terhadap kelas tingkat kemacetan berupa kata majemuk seperti “Macet Total”. Penelitian ini bertujuan untuk mengetahui kinerja klasifikasi tingkat kemacetan lalu lintas berbasis analisis *tweet* menggunakan algoritma LDA dengan data masukan berupa representasi kata *unigram*, *bigram*, dan *mixed-gram* (*unigram* + *bigram*). Kemudian, membandingkan kinerja klasifikasi berdasarkan setiap representasi kata tersebut. Hasil penelitian menunjukkan bahwa arsitektur terbaik dihasilkan oleh penerapan LDA sebagai data masukan berupa representasi kata *mixed-gram* dengan kombinasi parameter nilai *alpha* 0.1, *beta* 0.001, jumlah topik 10, dan jumlah iterasi 15000. Arsitektur tersebut, memiliki akurasi sebesar 100%. Oleh karena itu, untuk menghasilkan kinerja terbaik pembentukan model klasifikasi tingkat kemacetan lalu lintas berbasis analisis *tweet* dengan cara menerapkan LDA sebagai data masukan berupa *mixed-gram*.

**Kata Kunci** : *Latent Dirichlet Allocation*, *unigram*, *bigram*, *mixed-gram*, *Twitter*, kemacetan lalu lintas

## ABSTRACT

Traffic jam was main issue that often occurred in big cities and it gave a negative impact on environment around it. Twitter as one of the social media, could made users to share information about traffic jam nearby. Based on that issue, the utilization of Latent Dirichlet Allocation (LDA) could be used on tweet classification. Analysis from classification on traffic jam level based on tweet from twitter user were aimed for informing other people to know the locations of the traffic jam. While tweet was identified, bigram will be applied as an anticipation of declining performance on classification of traffic jam level class in form of compound word like “Macet Total”. The aim for this research was to find out the classification performance of traffic jam based on tweet analysis using LDA Algorithm with input data in form of unigram, bigram, and mixed-gram (unigram + bigram) word representation. Hereafter, the classification performance had been compared based on each representation of the word. The result of this research showed that the best architecture was produced by the application of LDA as an input data in form of mixed-gram word representation with parameter combination value alpha 0.1, beta 0.001, number of topics 10, and total amount iteration 15000. That architecture had 100% accuracy level. Therefore, in order to produce the best performance of building classification model of traffic jam level based on the analysis of the tweet, it should be performed by applying LDA to the data in the form of mixed-gram.

**Keywords :** *Latent Dirichlet Allocation, unigram, bigram, mixed-gram, Twitter, traffic jam*

## KATA PENGANTAR

Segala puji syukur bagi Tuhan Yang Maha Esa atas karunia-Nya yang diberikan kepada penulis sehingga penulis dapat menyelesaikan penulisan laporan tugas akhir yang berjudul “Klasifikasi Tingkat Kemacetan Lalu Lintas Berbasis Analisis *Tweet* Menggunakan *Latent Dirichlet Allocation* dan *N-Gram*”. Laporan tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu (S1) pada Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.

Dalam penyusunan laporan tugas akhir ini, penulis banyak mendapat bimbingan dan bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan rasa hormat dan terima kasih kepada:

1. Bapak Ragil Saputra, S.Si, M.Cs, selaku Ketua Departemen Ilmu Komputer/ Informatika.
2. Bapak Helmie Arif W., S.Si, M.Cs, selaku Koordinator Tugas Akhir Departemen Ilmu Komputer/ Informatika.
3. Ibu Dr. Retno Kusumaningrum, S.Si, M.Kom, selaku dosen pembimbing yang telah membantu dalam proses penelitian dan bimbingan hingga terselesaikannya laporan tugas akhir ini.
4. Bapak Rusdi Hariyanto dan Ibu Wiji Hastuti, selaku orang tua yang telah sabar dan memberikan dukungan dalam menyelesaikan tugas akhir ini.
5. Keluarga, Teman-teman dan Semua pihak yang telah membantu kelancaran dalam penyusunan tugas akhir ini, yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam laporan ini masih banyak kekurangan baik dari segi materi ataupun dalam penyajiannya karena keterbatasan kemampuan dan pengetahuan penulis. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan.

Semoga laporan tugas akhir ini dapat bermanfaat bagi pembaca pada umumnya dan penulis pada khususnya.

Semarang, 16 Desember 2016

Penulis,

Muhammad Ihsan Aji Wiedjayanto

24010312140050



## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PENGESAHAN .....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI .....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL .....	xv
DAFTAR LAMPIRAN .....	xviii
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan dan Manfaat.....	3
1.4. Ruang Lingkup .....	3
1.5. Sistematika Penulisan .....	4
BAB II TINJAUAN PUSTAKA .....	5
2.1. Perkembangan Penelitian Terkait Klasifikasi Tingkat Kemacetan Lalu Lintas Pada <i>Twitter</i> .....	5
2.2. Kemacetan .....	5
2.3. <i>Twitter</i> .....	6
2.4. Model <i>N-Gram</i> .....	6
2.5. <i>Latent Dirichlet Allocation</i> .....	7
2.6. Evaluasi.....	11
2.6.1. <i>K-fold Cross Validation</i> .....	11
2.6.2. <i>Confusion Matrix</i> .....	11
2.7. Pemrograman Berorientasi Objek.....	12
2.8. Pengembangan Perangkat Lunak.....	14
2.8.1. Fase Pada <i>Rational Unified Process</i> (RUP) .....	15
2.8.1.1. <i>Inception</i> (Permulaan) .....	15

2.8.1.2. <i>Elaboration</i> (Perluasan/Perencanaan) .....	16
2.8.1.3. <i>Construction</i> (Konstruksi) .....	17
2.8.1.4. <i>Transition</i> (Transisi) .....	17
2.8.2. <i>Workflows Pada Rational Unified Process (RUP)</i> .....	17
2.8.2.1. <i>Business Modeling</i> .....	18
2.8.2.2. <i>Requirements</i> .....	18
2.8.2.3. <i>Analysis &amp; Design</i> .....	18
2.8.2.4. <i>Implementation</i> .....	18
2.8.2.5. <i>Test</i> .....	19
2.8.2.6. <i>Deployment</i> .....	19
2.9. <i>Unified Modeling Language (UML)</i> .....	19
2.9.1. <i>Use Case Diagram</i> .....	20
2.9.2. <i>Activity Diagram</i> .....	22
2.9.3. <i>Class Diagram</i> .....	24
2.9.4. <i>Sequence Diagram</i> .....	26
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>28</b>
3.1. <i>Gambaran Umum Penelitian</i> .....	28
3.2. <i>Tahapan Penelitian</i> .....	30
3.2.1. <i>Pengumpulan Data</i> .....	31
3.2.2. <i>Preprocessing</i> .....	31
3.2.3.1. <i>Tokenisasi</i> .....	32
3.2.3.2. <i>Stopword Removal</i> .....	33
3.2.3.3. <i>Identifikasi Data Latih dan Data Uji</i> .....	33
3.2.3. <i>Pembentukan Model Klasifikasi</i> .....	34
3.2.3.1. <i>Pelatihan</i> .....	35
3.2.3.2. <i>Pengujian</i> .....	47
3.2.4. <i>Klasifikasi Tingkat Kemacetan Lalu Lintas Berbasis Analisis Tweet</i> <i>dengan Visualisasi Hasil Klasifikasi</i> .....	56
<b>BAB IV HASIL DAN ANALISIS</b> .....	<b>63</b>
4.1. <i>Data Penelitian</i> .....	63
4.1.1. <i>Data Tweet</i> .....	63
4.2. <i>Skenario Penelitian</i> .....	63
4.2.1. <i>Skenario 1</i> .....	64

4.2.2. Skenario 2 .....	65
4.2.3. Skenario 3 .....	66
4.2.4. Skenario 4 .....	67
4.3. Hasil Penelitian dan Analisis .....	67
4.3.1. Hasil Skenario 1 dan Analisis .....	67
4.3.2. Hasil Skenario 2 dan Analisis .....	69
4.3.3. Hasil Skenario 3 dan Analisis .....	71
4.3.4. Hasil Skenario 4 dan Analisis .....	73
4.4. Pengembangan Perangkat Lunak.....	84
4.4.1. Fase <i>Inception</i> .....	84
4.4.1.1. <i>Business Modeling</i> .....	84
4.4.1.2. <i>Requirements</i> .....	85
4.4.2. Fase <i>Elaboration</i> .....	88
4.4.2.1. Analisis .....	88
4.4.2.2. Desain .....	96
4.4.3. Fase <i>Construction</i> .....	105
4.4.3.1. Implementasi ( <i>Implementation</i> ).....	105
4.4.3.2. Pengujian ( <i>Testing</i> ).....	115
4.4.4. Fase <i>Transition</i> .....	116
BAB V PENUTUP .....	117
5.1. Kesimpulan .....	117
5.2. Saran .....	117
DAFTAR PUSTAKA.....	119
LAMPIRAN - LAMPIRAN .....	122

## DAFTAR GAMBAR

Gambar 2.1. LDA model (Kusumaningrum, et al., 2014).....	8
Gambar 2.2. Contoh LDA sebagai Generatif Model (Blei, 2012). ....	9
Gambar 2.3. Contoh Implementasi LDA sebagai inferensi (Blei, 2012). ....	9
Gambar 2.4. Visualiasai LDA sebagai inferensi (Kusumaningrum, 2013).....	10
Gambar 2.5. Model RUP ( <i>Rational Unified Process</i> ) (Rational Software Corporation, 1998) .....	15
Gambar 2.6. Contoh <i>Use Case Diagram</i> (Rumbaugh, et al., 2004).....	20
Gambar 2.7. Contoh Simbol <i>Use Case</i> (Sukamto & Shalahuddin, 2013).....	20
Gambar 2.8. Contoh Simbol <i>Actor</i> (Sukamto & Shalahuddin, 2013).....	21
Gambar 2.9. Contoh Simbol <i>Association</i> (Sukamto & Shalahuddin, 2013) .....	21
Gambar 2.10. Contoh Simbol <i>Extend</i> (Booch, et al., 1998).....	21
Gambar 2.11. Contoh Simbol <i>Generalization</i> (Sukamto & Shalahuddin, 2013).....	21
Gambar 2.12. Contoh Simbol <i>Include</i> (Booch, et al., 1998).....	22
Gambar 2.13. Contoh <i>Activity Diagram</i> untuk Proses Pesan Barang (Fowler, 2003) .....	22
Gambar 2.14. Contoh Simbol Status Awal (Sukamto & Shalahuddin, 2013).....	23
Gambar 2.15. Contoh Simbol Aktivitas (Rumbaugh, et al., 2004) .....	23
Gambar 2.16. Contoh Simbol <i>Decision</i> (Sukamto & Shalahuddin, 2013).....	23
Gambar 2.17. Contoh Simbol <i>Join</i> (Sukamto & Shalahuddin, 2013).....	23
Gambar 2.18. Contoh Simbol Status Akhir (Sukamto & Shalahuddin, 2013).....	24
Gambar 2.19. Contoh Simbol <i>Swimlane</i> Proses Pesan Barang (Booch, et al., 1998) .....	24
Gambar 2.20. Contoh <i>Class Diagram</i> (Booch, et al., 1998) .....	25
Gambar 2.21. Contoh <i>Sequence Diagram</i> (Booch, et al., 2005).....	26
Gambar 3.1. Gambaran Umum Penelitian .....	28
Gambar 3.2. Tahap Penelitian .....	31
Gambar 3.3. <i>Flowchart Preprocessing</i> .....	32
Gambar 3.4. <i>Flowchart Stopword Removal</i> .....	33
Gambar 3.5. <i>Flowchart</i> Pembentukan Model Klasifikasi .....	34
Gambar 3.6. <i>Flowchart</i> Proses Pelatihan .....	36
Gambar 3.7. <i>Flowchart</i> Inisialisasi Topik.....	38
Gambar 3.8. <i>Flowchart Generate Topic Random</i> .....	40
Gambar 3.9. <i>Flowchart Collapsed Gibbs Sampling</i> .....	40

Gambar 3.10. <i>Flowchart</i> Hitung PZD .....	43
Gambar 3.11. <i>Flowchart</i> Hitung PZC .....	44
Gambar 3.12. <i>Flowchart</i> Menghitung Nilai PZC.....	45
Gambar 3.13. <i>Flowchart</i> Hitung PZ.....	46
Gambar 3.14. <i>Flowchart</i> Hitung PWZ .....	47
Gambar 3.15. <i>Flowchart</i> Pengujian.....	48
Gambar 3.16. <i>Flowchart</i> Hitung PZD Uji.....	50
Gambar 3.17. <i>Flowchart</i> Hitung KLD .....	52
Gambar 3.18. <i>Flowchart</i> Hitung Akurasi.....	55
Gambar 3.19. Grafik Visualisasi Hasil Klasifikasi <i>Tweet</i> .....	57
Gambar 3.20. Visualisasi Klasifikasi Berdasarkan Kategori .....	57
Gambar 3.21. Visualisasi Isi <i>Tweet</i> .....	58
Gambar 3.22. <i>Flowchart</i> Proses Klasifikasi.....	58
Gambar 4.1. Skenario Penelitian .....	64
Gambar 4.2. Grafik Rata-Rata Hasil Akurasi Pengujian Data <i>Unigram</i> .....	68
Gambar 4.3. Grafik Arsitektur Terbaik <i>Collapsed Gibbs Sampling</i> Data <i>Unigram</i> .....	69
Gambar 4.4. Grafik Rata-Rata Hasil Akurasi Pengujian Data <i>Bigram</i> .....	70
Gambar 4.5. Grafik Hasil Akurasi Pengujian Data <i>Bigram</i> .....	71
Gambar 4.6. Grafik Rata-Rata Hasil Akurasi Pengujian Data <i>Mixed-Gram</i> .....	72
Gambar 4.7. Grafik Hasil Akurasi Pengujian Data <i>Mixed-Gram</i> .....	73
Gambar 4.8. Grafik Hasil Perbandingan Akurasi Pengujian Arsitektur Terbaik <i>Collapsed Gibbs Sampling</i> .....	74
Gambar 4.9. Grafik <i>Topic Proportion</i> Data <i>Unigram</i> .....	75
Gambar 4.10. Grafik <i>Topic Distribution</i> Data <i>Unigram</i> .....	77
Gambar 4.11. Grafik <i>Topic Proportion</i> Data <i>Bigram</i> .....	78
Gambar 4.12. Grafik <i>Topic Distribution</i> Data <i>Bigram</i> .....	80
Gambar 4.13. Grafik <i>Topic Proportion</i> Data <i>Mixed-Gram</i> .....	81
Gambar 4.14. Grafik <i>Topic Distribution</i> Data <i>Mixed-Gram</i> .....	83
Gambar 4.15. <i>Business Use Case Model</i> Klasifikasi Tingkat Kemacetan Lalu Lintas Berbasis Analisis <i>Tweet</i> .....	85
Gambar 4.16. <i>Use Case Diagram</i> Klasifikasi Tingkat Kemacetan Lalu Lintas Berbasis Analisis <i>Tweet</i> .....	86
Gambar 4.17. <i>Analysis Class Model</i> Menampilkan Informasi Cara Kerja dan	

Menyimpan <i>Stopword</i> .....	89
Gambar 4.18. <i>Analysis Class Model</i> Melakukan Klasifikasi <i>Tweet</i> .....	89
Gambar 4.19. <i>Analysis Class Model</i> Menampilkan Hasil Klasifikasi Tingkat Kemacetan	90
Gambar 4.20. <i>Analysis Class Model</i> Menampilkan Hasil Klasifikasi Sesuai Kelas .....	90
Gambar 4.21. <i>Analysis Class Model</i> Menampilkan Isi <i>Tweet</i> .....	91
Gambar 4.22. <i>Analysis Class Model</i> Menampilkan Informasi Tentang Aplikasi .....	91
Gambar 4.23. <i>Activity Diagram</i> Perangkat Lunak .....	93
Gambar 4.24. <i>Sequence Diagram</i> Menampilkan Informasi Cara Kerja dan Menyimpan <i>Stopword</i> .....	94
Gambar 4.25. <i>Sequence Diagram</i> Melakukan Klasifikasi <i>Tweet</i> .....	94
Gambar 4.26. <i>Sequence Diagram</i> Menampilkan Hasil Klasifikasi Tingkat Kemacetan ...	95
Gambar 4.27. <i>Sequence Diagram</i> Menampilkan Hasil Klasifikasi Sesuai Kelas .....	95
Gambar 4.28. <i>Sequence Diagram</i> Menampilkan Isi <i>Tweet</i> .....	96
Gambar 4.29. <i>Sequence Diagram</i> Menampilkan Informasi Tentang Aplikasi .....	96
Gambar 4.30. Gambar <i>Class Diagram</i> Klasifikasi Tingkat Kemacetan Lalu Lintas Berdasarkan Analisis <i>Tweet</i> .....	99
Gambar 4.31. Rancangan Antarmuka Halaman Informasi Cara Kerja 1 .....	99
Gambar 4.32. Rancangan Antarmuka Halaman Informasi Cara Kerja 2 .....	100
Gambar 4.33. Rancangan Antarmuka Halaman Informasi Cara Kerja 3 .....	100
Gambar 4.34. Rancangan Antarmuka Halaman Informasi Cara Kerja 4 .....	101
Gambar 4.35. Rancangan Antarmuka Halaman Informasi Cara Kerja 5 .....	101
Gambar 4.36. Rancangan Antarmuka Halaman Informasi Cara Kerja 6 .....	102
Gambar 4.37. Rancangan Antarmuka Halaman Utama Perangkat Lunak .....	102
Gambar 4.38. Rancangan Antarmuka Halaman Informasi Tentang Perangkat Lunak .....	103
Gambar 4.39. Rancangan Antarmuka Halaman Hasil Klasifikasi Tingkat Kemacetan....	103
Gambar 4.40. Rancangan Antarmuka Halaman Hasil Klasifikasi Berdasarkan Kelas .....	104
Gambar 4.41. Rancangan Antarmuka Halaman Isi <i>Tweet</i> .....	105
Gambar 4.42. Rancangan Antarmuka Pesan Peringatan Keluar dari Perangkat Lunak....	105
Gambar 4.43. Implementasi Antarmuka Halaman Informasi Cara Kerja 1 .....	107
Gambar 4.44. Implementasi Antarmuka Halaman Informasi Cara Kerja 2 .....	108
Gambar 4.45. Implementasi Antarmuka Halaman Informasi Cara Kerja 3 .....	109
Gambar 4.46. Implementasi Antarmuka Halaman Informasi Cara Kerja 4 .....	109
Gambar 4.47. Implementasi Antarmuka Halaman Informasi Cara Kerja 5 .....	110

Gambar 4.48. Implementasi Antarmuka Halaman Informasi Cara Kerja 6 .....	111
Gambar 4.49. Implementasi Antarmuka Halaman Utama Perangkat Lunak .....	111
Gambar 4.50. Implementasi Antarmuka Halaman Informasi Tentang Perangkat Lunak .	112
Gambar 4.51. Implementasi Antarmuka Halaman Hasil Klasifikasi Tingkat Kemacetan	113
Gambar 4.52. Implementasi Antarmuka Halaman Hasil Klasifikasi Berdasarkan Kelas .	113
Gambar 4.53. Implementasi Antarmuka Halaman Isi <i>Tweet</i> .....	114
Gambar 4.54. Implementasi Antarmuka Pesan Peringatan Keluar dari Perangkat Lunak	115
Gambar L.4.1. <i>Flowchart</i> Pembentukan <i>N-Gram</i> .....	137
Gambar L.4.2. <i>Flowchart</i> Pembentukan Kata (Konkatenasi) .....	137
Gambar L.4.3. <i>Flowchart</i> Proses Tokenisasi .....	138
Gambar L.5.1. <i>Flowchart</i> Identifikasi Data Latih dan Data Uji.....	140

## DAFTAR TABEL

Tabel 2.1. Perkembangan Penelitian Terkait Klasifikasi Tingkat Kemacetan Lalu Lintas Menggunakan Data <i>Tweet</i> .....	5
Tabel 2.2. Keterangan Notasi dan Definisi Model LDA.....	8
Tabel 2.3. Keterangan Notasi dan Definisi Persamaan 2.1, 2.2, 2.3 .....	10
Tabel 2.4. Model <i>Confusion Matrix</i> Klasifikasi Tingkat Kemacetan Lalu Lintas Berbasis Analisis <i>Tweet</i> .....	12
Tabel 2.5. Keterangan Notasi dan Definisi Persamaan <i>Accuracy</i> .....	12
Tabel 3.1. Pembagian Data Latih dan Data Uji.....	35
Tabel 3.2. Data Pelatihan.....	37
Tabel 3.3. <i>Vocabulary</i> Data Pelatihan .....	37
Tabel 3.4. Hasil Inisialisasi Topik .....	39
Tabel 3.5. Hasil Penghitungan PZD ( $\theta_{d,j}$ ).....	43
Tabel 3.6. Hasil Penghitungan PZC .....	45
Tabel 3.7. Data Pengujian.....	48
Tabel 3.8. <i>Vocabulary</i> Data Pengujian.....	49
Tabel 3.9. Daftar Jumlah Nilai PWZ Data Uji .....	51
Tabel 3.10. Daftar Nilai PZD Data Uji.....	52
Tabel 3.11. Daftar Nilai PZD Data Uji.....	53
Tabel 3.12. Daftar Nilai PZC Data Latih.....	53
Tabel 3.13. Hasil Perhitungan KLD Data Uji .....	54
Tabel 3.14. Daftar Kelas Aktual.....	55
Tabel 3.15. Daftar Perhitungan Akurasi.....	56
Tabel 3.16. Data <i>Tweet</i> Baru .....	59
Tabel 3.17. <i>Vocabulary</i> Data <i>Tweet</i> Baru.....	59
Tabel 3.18. Daftar Nilai PZD Data <i>Tweet</i> Baru .....	60
Tabel 3.19. Daftar Nilai PZC Latih .....	61
Tabel 3.20. Hasil Perhitungan DPQ Data <i>Tweet</i> Baru .....	61
Tabel 3.21. Hasil Perhitungan DQP Data <i>Tweet</i> Baru .....	62
Tabel 4.1. Skenario <i>Use Case</i> untuk <i>Use Case</i> Menampilkan Informasi Cara Kerja dan Menyimpan <i>Stopword</i> .....	86
Tabel 4.2. Skenario <i>Use Case</i> untuk <i>Use Case</i> Melakukan Klasifikasi <i>Tweet</i> .....	87



Tabel 4.3. Skenario <i>Use Case</i> untuk <i>Use Case</i> Menampilkan Hasil Klasifikasi Tingkat Kemacetan.....	87
Tabel 4.4. Skenario <i>Use Case</i> untuk <i>Use Case</i> Menampilkan Hasil Klasifikasi Sesuai Kelas.....	87
Tabel 4.5. Skenario <i>Use Case</i> untuk <i>Use Case</i> Menampilkan Isi <i>Tweet</i> .....	88
Tabel 4.6. Skenario <i>Use Case</i> untuk <i>Use Case</i> Menampilkan Informasi Tentang Aplikasi.....	88
Tabel 4.7. Identifikasi <i>Class</i> Analisis.....	91
Tabel 4.8. Tanggung Jawab dan Atribut <i>Class</i> .....	92
Tabel 4.9. Identifikasi <i>Class</i> Desain <i>Use Case</i> Menampilkan Informasi Cara Kerja dan Menyimpan <i>Stopword</i> .....	96
Tabel 4.10. Identifikasi <i>Class</i> Desain <i>Use Case</i> Melakukan Klasifikasi <i>Tweet</i> .....	97
Tabel 4.11. Identifikasi <i>Class</i> Desain <i>Use Case</i> Menampilkan Hasil Klasifikasi Tingkat Kemacetan.....	97
Tabel 4.12. Identifikasi <i>Class</i> Desain <i>Use Case</i> Menampilkan Hasil Klasifikasi Sesuai Kelas.....	97
Tabel 4.13. Identifikasi <i>Class</i> Desain <i>Use Case</i> Menampilkan Isi <i>Tweet</i> .....	98
Tabel 4.14. Identifikasi <i>Class</i> Desain <i>Use Case</i> Menampilkan Informasi Tentang Aplikasi.....	98
Tabel 4.15. Daftar <i>Class</i> Desain.....	98
Tabel 4.16. <i>Class</i> Implementasi .....	106
Tabel L.3.1. Data <i>Tweet</i> Penelitian.....	127
Tabel L.6.1. Hasil Pembagian Data Latih .....	141
Tabel L.6.2. Hasil Pembagian Data Uji.....	141
Tabel L.7.1. Kombinasi Parameter Penelitian.....	143
Tabel L.8.1. Hasil Probabilitas $p(z_i = j z_{-i}, w)$ Data Latih .....	144
Tabel L.8.2. Hasil Normalisasi Nilai $p(z_i = j z_{-i}, w)$ Data Latih.....	145
Tabel L.8.3. Hasil <i>Sampling</i> Topik Data Latih .....	147
Tabel L.8.4. Hasil Penghitungan PWZ Data Latih.....	148
Tabel L.8.5. Daftar Nilai PWZ Data Uji .....	149
Tabel L.8.6. Hasil Penghitungan DPQ Data Uji .....	150
Tabel L.8.7. Hasil Perhitungan DQP Data Uji .....	150
Tabel L.8.8. Daftar Nilai KLD Data Uji.....	151

Tabel L.9.1. Daftar Nilai PWZ Data <i>Tweet</i> Baru .....	152
Tabel L.9.2. Hasil Penghitungan KLD Data <i>Tweet</i> Baru .....	152
Tabel L.10.1. Hasil Pengujian terhadap Data dengan Representasi Kata <i>Unigram</i> .....	153
Tabel L.10.2. Hasil Pengujian terhadap Data dengan Representasi Kata <i>Bigram</i> .....	154
Tabel L.10.3. Hasil Pengujian terhadap Data dengan Representasi Kata <i>Mixed-Gram</i> ( <i>Unigram + Bigram</i> ) .....	155
Tabel L.11.1. Hasil Analisis Perbandingan Akurasi Pengujian Arsitektur Terbaik <i>Collapsed Gibbs Sampling</i> .....	156
Tabel L.11.2. Hasil Probabilitas Kata-Topik (PWZ) Terbaik Data <i>Unigram</i> .....	156
Tabel L.11.3. Hasil Probabilitas Kata-Topik (PWZ) Terbaik Data <i>Bigram</i> .....	158
Tabel L.11.4. Hasil Probabilitas Kata-Topik (PWZ) Terbaik Data <i>Mixed-Gram</i> .....	160
Tabel L.12.1. Daftar Rencana Pengujian Perangkat Lunak Klasifikasi Tingkat Kemacetan Lalu Lintas Berbasis Analisis <i>Tweet</i> .....	163
Tabel L.12.2. Daftar Hasil Pengujian Perangkat Lunak Klasifikasi Tingkat Kemacetan Lalu Lintas Berbasis Analisis <i>Tweet</i> .....	163

## DAFTAR LAMPIRAN

Lampiran 1. Kartu Bimbingan Tugas Akhir.....	124
Lampiran 2. Kartu Keikutsertaan Seminar TA 1.....	126
Lampiran 3. Data Penelitian .....	127
Lampiran 4. Proses Tokenisasi.....	137
Lampiran 5. Proses Identifikasi Data Latih dan Data Uji.....	139
Lampiran 6. Identifikasi Data Latih dan Data Uji .....	141
Lampiran 7. Kombinasi Parameter.....	143
Lampiran 8. Pembentukan Model Klasifikasi .....	144
Lampiran 9. Klasifikasi Tingkat Kemacetan Lalu Lintas Berbasis Analisis Tweet dengan Visualisasi Hasil Klasifikasi .....	152
Lampiran 10. Hasil Pembentukan Model Klasifikasi.....	153
Lampiran 11. Analisis Perbandingan Akurasi Pengujian Arsitektur Terbaik <i>Collapsed Gibbs Sampling</i> .....	156
Lampiran 12. Daftar Rencana Pengujian dan Hasil Pengujian .....	163

# BAB I

## PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan dalam penelitian tugas akhir mengenai klasifikasi tingkat kemacetan lalu lintas berbasis analisis *tweet* menggunakan *Latent Dirichlet Allocation* dan *n-gram*.

### 1.1. Latar Belakang

Perkembangan teknologi informasi dan komunikasi saat ini telah berkembang pesat, terutama bidang komunikasi. Teknologi terhadap akses komunikasi yang semakin maju memudahkan masyarakat dalam bersosialisasi. Kemudahan tersebut ditawarkan melalui munculnya banyak media sosial yang dapat digunakan oleh semua orang untuk berinteraksi atau sekedar mencari hiburan melalui dunia maya contohnya, *Facebook* dan *Twitter*. *Twitter* menawarkan jaringan sosial berupa *microblog*. Disebut *microblog* karena situs ini memungkinkan penggunanya mengirim dan membaca pesan blog seperti pada umumnya namun terbatas hanya sejumlah 140 karakter yang ditampilkan pada halaman profil pengguna. Pesan dalam *Twitter* dikenal dengan sebutan *tweet* (Faradhillah, et al., 2016).

Di Indonesia sendiri, selama tahun 2014 jumlah pengguna internet telah mencapai angka 88,1 juta jiwa pengguna dari total jumlah penduduk 252,4 juta jiwa (Asosiasi Penyelenggara Jasa Internet Indonesia, 2014). Hal itu menunjukkan bahwa masyarakat Indonesia semakin sering menggunakan internet. Salah satu media yang ditawarkan oleh internet dan intens digunakan oleh masyarakat Indonesia adalah media sosial, seperti *Twitter*. Indonesia menjadi negara ke-lima teraktif dan Jakarta menjadi kota nomor satu yang teraktif memposting di *Twitter* (SemioCast, 2012). Melalui *Twitter* setiap orang dapat berbagi informasi secara *real-time*, saling terhubung dengan pengguna lain di mana pun mereka berada, hingga mengekspresikan beragam pendapat tentang apa yang sedang terjadi di sekitar mereka contohnya, berbagi pesan mengenai keadaan lalu-lintas terkini yang mereka alami.

Sebagai salah satu negara berkembang, Indonesia seperti negara berkembang lainnya yang mengalami permasalahan-permasalahan lebih kompleks dibandingkan

dengan negara-negara maju, mulai dari pertumbuhan penduduk yang tinggi, kesenjangan sosial, kurangnya sarana dan prasarana yang menunjang pembangunan itu sendiri, serta kemacetan.

Permasalahan lalu lintas merupakan suatu masalah sulit yang harus dipecahkan bersama dan sangat penting untuk segera diselesaikan. Apabila masalah lalu lintas tidak terpecahkan, maka semua kerugian yang timbul akibat masalah ini akan ditanggung oleh masyarakat itu sendiri, dan apabila masalah ini dapat terpecahkan dengan baik, maka masyarakat sendiri yang akan mendapatkan manfaatnya.

Penelitian mengenai Klasifikasi Tingkat Kemacetan Lalu Lintas berbasis *tweet* telah banyak dikembangkan dengan menggunakan berbagai metode seperti *Naive Bayesian Classification* (Rodiansyah & Winarko, 2012) dan *Support Vector Machine* (Susilowati, et al., 2015). Pengguna *Naive Bayesian Classification* hanya menitik beratkan pada teorema *Bayes* yang mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Sedangkan, *Support Vector Machine* memiliki kelebihan karena mampu mengklasifikasikan data berdimensi tinggi yang dalam konteks penelitian tersebut adalah data berupa teks.

Akan tetapi, penggunaan *Naive Bayesian Classification* dan *Support Vector Machine* memberikan kemungkinan pengklasifikasian dokumen secara kaku, yaitu sebuah data hanya termasuk satu kelas sehingga akan berpengaruh terhadap kinerja klasifikasi, sedangkan untuk mendapatkan kinerja klasifikasi yang lebih baik dapat dilakukan dengan mengetahui topik dari masing-masing dokumen. Sehingga sebuah data dapat dikategorikan ke dalam beberapa kelas sesuai dengan tingkat kemiripan dari topik-topiknya. Hal tersebut dapat diselesaikan menggunakan *Probabilistic Topic Modeling* (Blei, 2012).

Kedua penelitian tersebut hanya mengklasifikasikan *tweet* berdasarkan kata sebagai unit terkecil sehingga pada kasus mengidentifikasi tingkat kemacetan lalu lintas akan mengalami penurunan kinerja untuk kelas berupa kata majemuk seperti ramai lancar, padat merayap dan macet total. Salah satu teknik yang dapat diterapkan adalah penggunaan *Bigram* (Hemalatha, et al., 2012).

*Probabilistic Topic Models* merupakan metode dalam *Topic Modeling* yang menyediakan kemudahan untuk mengorganisir dan meringkas arsip elektronik dalam skala yang besar (Blei, 2012). Salah satu bentuk *Probabilistic Topic Models* yang sangat populer digunakan, dan memiliki asumsi yang lebih komprehensif pada

pembangkitan teks dibandingkan model lainnya adalah *Latent Dirichlet Allocation* (Liu, 2013).

## 1.2. Rumusan Masalah

Berdasarkan pada uraian latar belakang di atas, maka didapat suatu rumusan masalah tentang bagaimana menerapkan *Latent Dirichlet Allocation* dan *N-Gram* untuk Klasifikasi Tingkat Kemacetan Lalu Lintas Berbasis Analisis *Tweet*.

## 1.3. Tujuan dan Manfaat

Tujuan yang ingin dicapai dalam penelitian tugas akhir ini adalah sebagai berikut:

1. Mengetahui kinerja penerapan *Latent Dirichlet Allocation* dan *Unigram* untuk klasifikasi tingkat kemacetan lalu lintas berbasis analisis *tweet*.
2. Mengetahui kinerja penerapan *Latent Dirichlet Allocation* dan *Bigram* untuk klasifikasi tingkat kemacetan lalu lintas berbasis analisis *tweet*.
3. Mengetahui kinerja penerapan *Latent Dirichlet Allocation* dan *Mixed Gram (Unigram + Bigram)* untuk klasifikasi tingkat kemacetan lalu lintas berbasis analisis *tweet*.
4. Membandingkan kinerja penerapan *Unigram*, *Bigram* dan *Mixed Gram (Unigram + Bigram)* pada klasifikasi tingkat kemacetan lalu lintas berbasis analisis *tweet*.

Adapun manfaat yang diharapkan dari penelitian tugas akhir ini adalah sebagai berikut:

1. Informasi yang bermanfaat bagi pemerintah atau masyarakat umum mengenai kondisi tingkat kemacetan lalu lintas pada suatu daerah tertentu.
2. Solusi alternatif bagi pemerintah sebagai informasi pendukung dalam pengambilan keputusan untuk melakukan pengaturan khusus terhadap lalu lintas pada daerah yang sering terjadi kemacetan.

## 1.4. Ruang Lingkup

Dalam penyusunan penelitian ini, diberikan ruang lingkup yang jelas agar pembahasan lebih terarah dan tidak menyimpang dari tujuan penulisan. Adapun ruang lingkup dalam penelitian ini adalah sebagai berikut :

1. Data *tweet* yang digunakan adalah *tweet* berbahasa Indonesia.

2. Data yang diambil hanya berupa teks, tidak termasuk suara, gambar maupun video.
3. Proses pembelajaran dilakukan secara *offline*.
4. Representasi data *n-gram* yang diujicobakan hanya *unigram*, *bigram*, dan *mixed gram (unigram + bigram)*.
5. Penentuan klasifikasi tingkat kemacetan lalu lintas berupa lancar, ramai lancar, padat merayap, macet total, dan tidak berhubungan dengan keadaan lalu lintas.

## 1.5. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam tugas akhir ini terbagi menjadi beberapa pokok bahasan, yaitu:

### BAB I PENDAHULUAN

Bab ini memberikan gambaran tentang latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup serta sistematika penulisan.

### BAB II TINJAUAN PUSTAKA

Bab ini membahas mengenai sejumlah kajian pustaka yang berhubungan dengan penelitian tugas akhir ini. Kajian tersebut meliputi penjelasan mengenai perkembangan penelitian terkait klasifikasi tingkat kemacetan lalu lintas pada *Twitter*, kemacetan, *Twitter*, model *n-gram*, *Latent Dirichlet Allocation*, evaluasi, pemrograman berorientasi objek, pengembangan perangkat lunak, dan *Unified Modeling Language*.

### BAB III METODOLOGI PENELITIAN

Bab ini membahas mengenai langkah-langkah yang dilakukan pada penelitian tugas akhir. Pada bab ini diawali dengan gambaran umum penelitian kemudian dilanjutkan dengan tahapan penelitian yang berisi pengumpulan data, *preprocessing*, pembentukan model klasifikasi tingkat kemacetan lalu lintas, dan proses klasifikasi dengan visualisasi.

### BAB IV HASIL DAN ANALISIS

Bab ini membahas mengenai data penelitian, skenario penelitian, hasil dan analisis penelitian, serta pengembangan perangkat lunak.

### BAB V KESIMPULAN

Bab ini membahas mengenai kesimpulan dari uraian yang telah dijabarkan pada bab-bab sebelumnya dan saran untuk pengembangan penelitian lebih lanjut.