



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

**VISUALIZATION OF ROAD ACCIDENT DATA  
AT JABODETABEK AREA  
IN YEAR 2014 - 2017**

FINAL PROJECT



Submitted to Fulfilled the Requirements of Obtaining a Computer Degree  
(Bachelor of Computer)

**Sudecanto**  
**14110310008**

**INFORMATION SYSTEM STUDY PROGRAM  
FACULTY OF ENGINEERING AND INFORMATICS  
MULTIMEDIA NUSANTARA UNIVERSITY  
TANGERANG  
2018**

**VISUALISASI DATA TINGKAT KECELAKAAN  
LALU LINTAS DI JABODETABEK  
TAHUN 2014 - 2017**

SKRIPSI



Diajukan Guna Memenuhi Persyaratan Memperoleh  
Gelar Sarjana Komputer (S.Kom.)

**Sudecanto**

**14110310008**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2018**

## PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi ini adalah karya ilmiah saya sendiri, bukan plagiat dari karya ilmiah yang ditulis oleh orang lain atau lembaga lain, dan semua karya ilmiah orang lain atau lembaga lain yang dirujuk dalam skripsi ini telah disebutkan sumber kutipannya serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan / penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah skripsi yang telah saya tempuh dan status keserjanaan strata satu yang sudah diterima akan dicabut.

Tangerang, 2 Mei 2018



Sudecanto

## HALAMAN PERSETUJUAN

Skripsi dengan judul

“Visualisasi Data Tingkat

Kecelakaan Lalu Lintas di Jabodetabek

Tahun 2014 - 2017”

oleh

Sudecanto

telah disetujui untuk diajukan pada

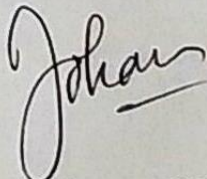
Sidang Ujian Skripsi Universitas Multimedia Nusantara

Tangerang, 2 Mei 2018

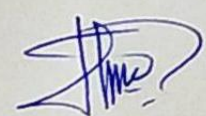
Menyetujui,

**Dosen Pembimbing**

**Ketua Program Studi**



Johan Setiawan, S.Kom., M.M, M.B.A

 2/5/18

Ririn Ikana Desanti, S.Kom., M.Kom.

## HALAMAN PENGESAHAN

Skripsi dengan judul


“Visualisasi Data Tingkat Kecelakaan Lalu Lintas  
di Jabodetabek Tahun 2014 - 2017”

oleh

Sudecanto

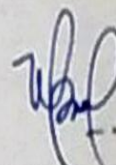
telah diuji pada hari Senin, 23 April 2018  
pukul 10.30 s.d 12.00 dan dinyatakan lulus  
dengan susunan penguji sebagai berikut.

**Ketua Sidang**

 2/5/18

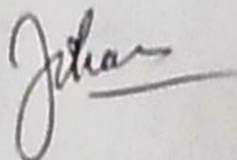
Marcelli Indriana, S.Kom., M.Sc.

**Penguji**

 2/5/18

Wella, S.Kom., M.MSI.


**Dosen Pembimbing**



Johan Setiawan, S.Kom., M.M., M.B.A.

**Disahkan oleh**

**Ketua Program Studi Sistem Informasi – UMN**

 2/5/18

Ririn Ikana Desanti, S.Kom., M.Kom.

## KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Kuasa sehingga skripsi yang berjudul “Visualisasi Data Tingkat Kecelakaan Lalu Lintas di Jabodetabek” dapat selesai tepat pada waktunya. Skripsi ini penulis ajukan kepada Program Strata 1, Program Studi Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi, Universitas Multimedia Nusantara.

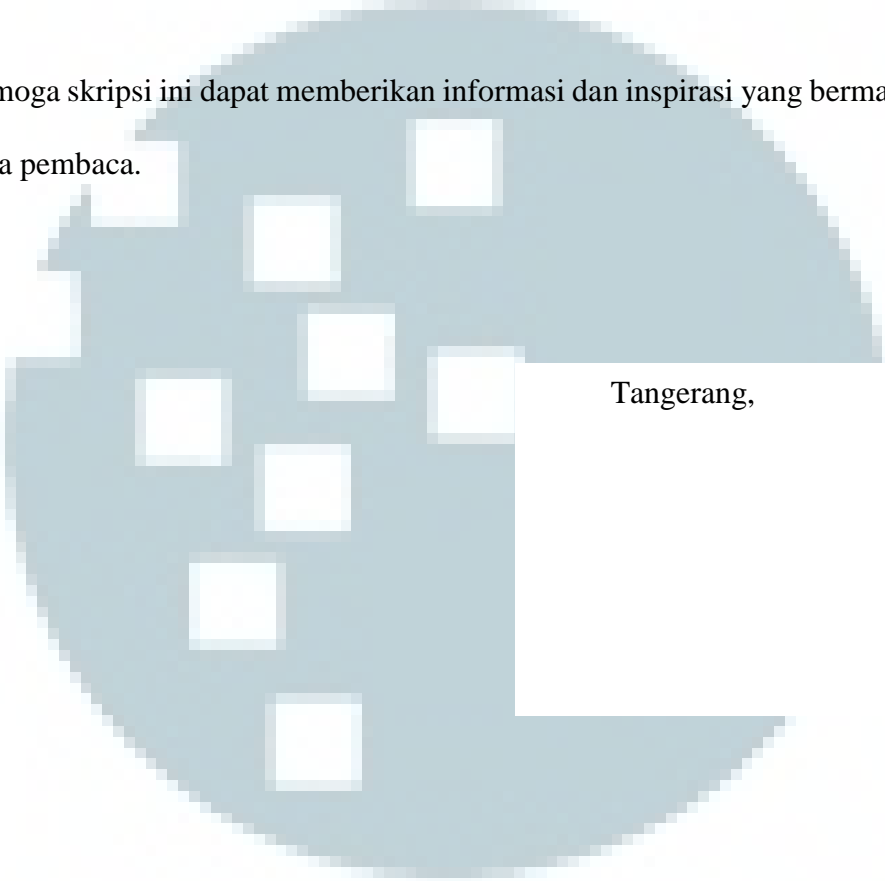
Dengan berakhirnya proses penulisan skripsi ini, penulis ingin mengucapkan terima kasih kepada Universitas Multimedia Nusantara yang telah memberi beasiswa kepada penulis selama menempuh pendidikan di Universitas Multimedia Nusantara sehingga dapat membantu meringankan penulis dalam membayar biaya kuliah penulis hingga selesai.

Selain itu, penulis juga ingin mengucapkan terima kasih kepada:

1. Bapak Johan Setiawan, S.Kom., M.M, M.B.A yang telah memberikan bimbingan dan saran-saran yang diberikan kepada penulis selama pengerjaan skripsi
2. Ibu Friska Natalia, Ph.D yang telah membimbing proposal skripsi dalam mata kuliah metode research sistem informasi.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom. Selaku Ketua Prodi Sistem Informasi

Tidak lupa penulis mengucapkan terima kasih kepada orang tua dan keluarga yang telah memberikan semangat dan doa kepada penulis dalam penyelesaian skripsi ini.

Semoga skripsi ini dapat memberikan informasi dan inspirasi yang bermanfaat bagi para pembaca.



Tangerang,

Penulis

UMN



## DAFTAR ISI

PERNYATAAN.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR .....	iv
DAFTAR ISI.....	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR GRAFIK.....	xii
DAFTAR RUMUS .....	xiii
ABSTRAK.....	xiv
<i>ABSTRACT</i> .....	xv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	4
1.7 Jadwal Penelitian.....	6
BAB II LANDASAN TEORI.....	7
2.1 <i>Text mining</i> .....	7

2.2	<i>Big Data</i> .....	7
2.3	Visualisasi Data .....	8
2.4	<i>Cross-Industry Standard Process for Data mining (CRISP-DM)</i> .....	9
2.5	Tahap <i>Data mining</i> .....	11
2.6	<i>Naive Bayes Classifier</i> .....	13
2.7	<i>K-means Clustering</i> .....	14
2.8	<i>Tools</i> .....	16
2.9	Grafik.....	21
BAB III METODOLOGI PENELITIAN.....		29
3.1	Objek Penelitian .....	29
3.2	Penelitian Terdahulu.....	29
3.3	Perbandingan Metode .....	33
3.4	Perbandingan <i>Tools</i> .....	36
3.5	Metode <i>CRISP-DM</i> .....	37
3.5.1	<i>Business Understanding</i> .....	37
3.5.2	<i>Data Understanding</i> .....	37
3.5.3	<i>Data Preparation</i> .....	38
3.5.4	<i>Modelling</i> .....	38
3.5.5	<i>Evaluation</i> .....	38
3.5.6	<i>Deployment</i> .....	38
BAB IV ANALISIS DAN HASIL PENELITIAN .....		39
4.1	<i>Business Understanding</i> .....	39

4.2	<i>Data Understanding</i> .....	40
4.3	<i>Data Preparation</i> .....	41
4.4	<i>Modelling</i> .....	61
4.5	<i>Evaluation</i> .....	89
4.6	<i>Deployment</i> .....	89
BAB V KESIMPULAN DAN SARAN.....		91
5.1	Kesimpulan.....	91
5.2	Saran.....	92
DAFTAR PUSTAKA .....		94
LAMPIRAN.....		96

UMMN

## DAFTAR TABEL

Tabel 1.1 Jadwal Penelitian.....	6
Tabel 3.2 Penelitian Terdahulu .....	29
Tabel 3.3 <i>Road Accident Attributes</i> .....	31
Tabel 3.4 Perbandingan Metode Visualisasi Data .....	33
Tabel 3.5 Perbandingan Algoritma .....	34
Tabel 4.6 <i>Twitter Username</i> “Kecelakaan” .....	45
Tabel 4.7 <i>Command Python</i> TMC Polda Metro Jaya 2017 .....	46
Tabel 4.8 <i>Command Python</i> TMC Polda Metro Jaya 2016 .....	47
Tabel 4.9 <i>Command Python</i> TMC Polda Metro Jaya 2015 .....	48
Tabel 4.10 <i>Command Python</i> TMC Polda Metro Jaya 2014 .....	48
Tabel 4.11 Hasil Data <i>Tweets</i> .....	49
Tabel 4.12 <i>Cleansing Result</i> Tahap 1 .....	50
Tabel 4.13 <i>Cleansing Criteria</i> Tahap 1 .....	50
Tabel 4.14 <i>Cleansing Result</i> Tahap 2 & 3 .....	59
Tabel 4.15 Kecelakaan Terbanyak .....	65
Tabel 4.16 <i>Cluster</i> 1 Tangerang.....	75
Tabel 4.17 <i>Cluster</i> 2 Bekasi .....	76
Tabel 4.18 <i>Cluster</i> 3 Bogor.....	78
Tabel 4.19 <i>Cluster</i> 4 Jakarta.....	79
Tabel 4.20 <i>Cluster</i> 5 Depok .....	81
Tabel 4.21 <i>Cluster</i> Kesimpulan Jenis Kendaraan dengan Musim .....	82
Tabel 4.22 <i>Cluster</i> 1 T4 .....	84

Tabel 4.23 <i>Cluster 2</i> T6 .....	85
Tabel 4.24 <i>Cluster 3</i> T3 .....	85
Tabel 4.25 <i>Cluster 4</i> T2 .....	86
Tabel 4.26 <i>Cluster 5</i> T5 .....	87
Tabel 4.27 <i>Cluster 6</i> T1 .....	88
Tabel 4.28 Kesimpulan Jenis Kendaraan Berdasarkan Waktu .....	88



## DAFTAR GAMBAR

Gambar 2.1 <i>CRISP-DM Cycle</i> .....	9
Gambar 2.2 <i>Python</i> .....	16
Gambar 2.3 <i>PHP</i> .....	17
Gambar 2.4 <i>Tweepy</i> .....	18
Gambar 2.5 <i>Power BI</i> .....	19
Gambar 2.6 <i>R Script</i> .....	20
Gambar 2.7 <i>Line Chart</i> .....	21
Gambar 2.8 <i>Pie Chart</i> .....	22
Gambar 2.9 <i>Gauge Chart</i> .....	23
Gambar 2.10 <i>Power BI Tabel</i> .....	23
Gambar 2.11 <i>Power BI Slicer</i> .....	24
Gambar 2.12 <i>Donut Chart</i> .....	25
Gambar 2.13 <i>Clustered Column Chart</i> .....	26
Gambar 2.14 <i>Map Chart</i> .....	27
Gambar 2.15 <i>Cluster Chart</i> .....	28
Gambar 3.16 <i>Road Accident Map Chart</i> .....	32
Gambar 3.17 <i>Power BI Comparison</i> .....	36
Gambar 4.18 <i>Direktori Python Master</i> .....	43
Gambar 4.19 <i>Get Old Tweets</i> .....	44
Gambar 4.20 <i>PHP Auto Generate Column Data</i> .....	52
Gambar 4.21 <i>PHP Shortened text</i> .....	58

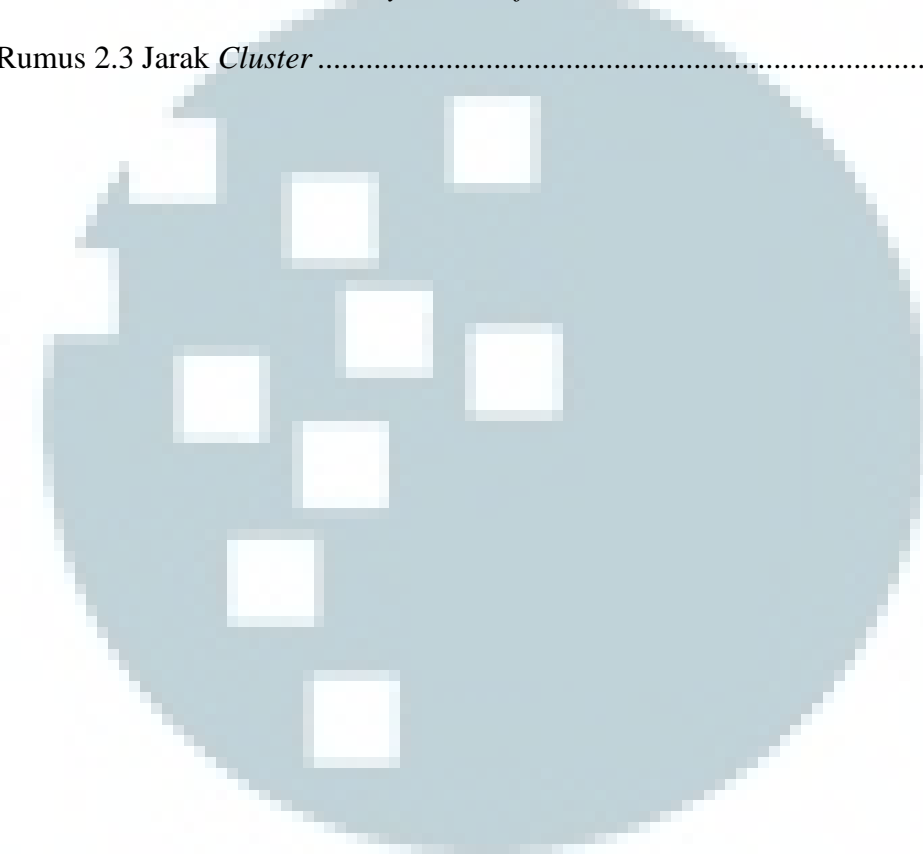
## DAFTAR GRAFIK

Grafik 4.1 Dashboard 1 .....	62
Grafik 4.2 Jumlah Kecelakaan Berdasarkan Kota .....	63
Grafik 4.3 Kecelakaan Berdasarkan Jenis Kendaraan .....	64
Grafik 4.4 Lokasi Rawan Kecelakaan <i>Power BI</i> .....	66
Grafik 4.5 Grafik <i>Twitter User Preference</i> .....	67
Grafik 4.6 Jumlah Kecelakaan Berdasarkan Waktu .....	68
Grafik 4.7 Jumlah Kecelakaan Berdasarkan <i>Interval</i> Waktu 4 jam.....	69
Grafik 4.8 Grafik Kecelakaan Berdasarkan Musim.....	70
Grafik 4.9 Kecelakaan Berdsarkan Merk Kendaraan .....	71
Grafik 4.10 Grafik <i>Cluster K-means</i> Tahap Pertama.....	73
Grafik 4.11 Grafik <i>Cluster K-means</i> – 1 (3 Variabel) .....	74
Grafik 4.12 Grafik <i>Cluster K-means</i> – 2 ( 2 Variabel) .....	83

UMMN

## DAFTAR RUMUS

Rumus 2.1 <i>Naive Bayes Classifier</i> .....	13
Rumus 2.2 <i>Posterior Naive Bayes Classifier</i> .....	14
Rumus 2.3 <i>Jarak Cluster</i> .....	15



UMN



# VISUALISASI DATA TINGKAT KECELAKAAN LALU LINTAS DI JABODETABEK TAHUN 2014 - 2017

## ABSTRAK

Oleh: Sudecanto

Penelitian ini bertujuan untuk melakukan visualisasi data untuk merepresentasikan data kecelakaan lalu lintas yang terjadi di area Jabodetabek pada tahun 2014-2017. Kecelakaan lalu lintas merupakan masalah bagi seluruh negara karena menyebabkan kerugian berupa materi dan jiwa. Sulitnya untuk mendapatkan data kecelakaan lalu lintas di area Jabodetabek membuat masyarakat tidak mengetahui tingkat kecelakaan lalu lintas di area Jabodetabek.

Data kecelakaan lalu lintas diperoleh dengan melakukan *text mining* yang didapat dari *social media Twitter* menggunakan *tweepy api* dengan menarik *tweets*. *tweets* yang diambil adalah *tweet* yang memiliki kata “kecelakaan” pada periode waktu Januari 2014 hingga Desember 2017.

Data diproses dan dilabel secara otomatis dengan menggunakan program *PHP* yang telah dibuat dengan algoritma yang membaca pola sebuah teks dan menemukan kata kunci yang dicari untuk disimpan ke kolom baru berupa data yang diinginkan. Data yang sudah dilabel dalam bentuk tabel diproses dengan *tools Power BI*.

Data yang sudah diolah diproses dengan *tools Power BI* memberikan *output* data visual yang memberikan informasi detail tentang data kecelakaan lalu lintas. Data ini akan di kluster dengan metode *k-means* dengan hasil analisa mobil dan truk merupakan jenis kendaraan yang paling sering mengalami kecelakaan, kecelakaan lebih sering terjadi pada musim hujan, dan waktu rawan kecelakaan adalah T2 yaitu jam 04.00 WIB s.d 08.00 WIB.

**Kata kunci:** *text mining, k-means, data visual.*

**VISUALIZATION OF ROAD ACCIDENTS  
AT JABODETABEK  
YEAR 2014 - 2017**

**ABSTRACT**

By: Sudecanto

*This study aims to give data visualization to represent road accidents data that occurred in Jabodetabek during year 2014-2017. Road accidents is a problem for all countries across the world because it causes material and life losses. It is very hard to get information about road accidents data in Jabodetabek make people do not know the threat of road accidents in Jabodetabek.*

*Road Accidents data will be obtained by doing text mining from social media Twitter using tweepy API, tweets that will be taken is a tweet that has the word "accident" during period January 2014 to December 2017.*

*Data is processed and labeled automatically by using a PHP program which has been created with an algorithm that reads a text pattern and finds the keyword, the keyword information will be stored to a new column for desired information. Data that has been labeled in structured table form is processed with Power BI tools.*

*Data will be processed using Power BI tools to provide visualization data output that provides detailed information about road accidents that occurred in Jabodetabek during year 2014-2017. Data will be clustered using k-means algorithm which result information are Cars and trucks are the type of vehicle that most often have accidents, accidents are more common in rainy season, and accidents prone time is T2 (04.00 a.m to 08.00 a.m).*

**Keywords :** *text mining, k-means, data visual.*

U M N