



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.

BAB III

METODOLOGI PENELITIAN

3.1 Objek Penelitian

Objek dari penelitian ini adalah kecelakaan lalu lintas yang berasal dari *tweets* yang didapat melalui *social media Twitter*. *Tweets* yang diambil harus memenuhi syarat yaitu mengandung kata kunci “Kecelakaan” dan berasal dari *username* “TMCPoldaMetroJaya” dengan periode waktu yaitu Januari 2014 s.d Desember 2017. *Tweets* yang diambil juga hanya *tweets* dengan lokasi kecelakaan di kota Jakarta, Bogor, Depok, Tangerang, dan Bekasi.

3.2 Penelitian Terdahulu

Tabel 3.2 Penelitian Terdahulu

No	Peneliti	Tahun Penelitian	Judul Penelitian	Masalah Penelitian	Hasil Penelitian	Penerapan Penelitian
1	Sachin Kumar dan Durga Toshniwara	2015	“A data mining framework to analyze road accident data”	Tujuan utama dalam menganalisa data kecelakaan lalu lintas adalah mengidentifikasi penyebab utama yang menyebabkan terjadi kecelakaan lalu lintas.	Tujuan utama dalam menganalisa data kecelakaan lalu lintas adalah mengidentifikasi penyebab utama yang menyebabkan terjadinya kecelakaan lalu lintas.	Penelitian ini dijadikan pedoman dalam membuat analisa <i>cluster</i>

Tabel 3.2 Penelitian Terdahulu (Lanjutan)

No	Peneliti	Tahun Penelitian	Judul Penelitian	Masalah Penelitian	Hasil Penelitian	Penerapan Penelitian
2	Luis Ramos	2015	“Detection of road accident accumulation zones with a visual analytic approach”	Pada zaman ini kecelakaan lalu lintas merupakan masalah utama permasalahan kesehatan maka perlunya ada prediksi terjadinya kecelakaan lalu lintas. Kecelakaan lalu lintas di dunia terjadi sebanyak 1.200.000 orang meninggal di seluruh dunia.	1. Titik terjadinya kecelakaan. 2. Data visual konsentrasi terjadinya kecelakaan / <i>gray spot</i>	Penelitian ini dijadikan pedoman dalam membuat <i>mapping visual</i> lokasi rawan kecelakaan
3	Rafal S. Jurecki dan Marek Jaśkiewicz	2012	“Analysis of road accidents in Poland over the last ten years”	Perlunya analisa angka kecelakaan lalu lintas, korban luka, korban jiwa pada tahun-tahun akhir ini di polandia di berbagai musim termasuk jenis kendaraan yang digunakan korban dan penyebab terjadinya kecelakaan lalu lintas.	1. Data visual jumlah kecelakaan lalu lintas dan korban luka/jiwa di polandia 2. Analisa <i>trend</i> kecelakaan pengemudi mobil 3. Data visual jumlah kecelakaan per bulan, jumlah korban jiwa kecelakaan lalu lintas 4. Analisa penyebab terjadinya kecelakaan lalu lintas	Penelitian ini dijadikan pedoman dalam pembuatan data visual <i>dashboard</i> kecelakaan lalu lintas di Jabodetabek tahun 2014-2017

Tabel 3.2 penelitian terdahulu menjelaskan penelitian dengan judul “*A data mining framework to analyze road accident data*” yang diterapkan dalam penelitian ini sebagai pedoman bagaimana cara melakukan pengelompokan data dengan metode *clustering*. Algoritma yang digunakan dalam penelitian ini juga sama yaitu *k-means clustering*.

Tabel 3.3 Road Accident Attributes

Sumber : (Kumar, 2015)

Table 1 Road accident attributes

S. no.	Attribute	Code	Value	Total	Criticality	
					Critical	Non-critical
1	NOV: number of injury	1	1 injury	5932	689	5243
		2	2 injuries	2598	451	2147
		+2	>2 injuries	3044	114	2930
2	AGE: age	CHL	< 18 years	988	268	720
		YNG	18–30 years	5954	654	5300
		ADL	30–60 years	3045	165	2880
		SNR	>60 years	1587	167	1420
3	GND: gender	M	Male	8625	952	7673
		F	Female	2949	302	2647
4	TOD: time of day	T1	[0–4]	678	45	633
		T2	[4–8]	1032	164	868
		T3	[8–12]	1358	258	1100
		T4	[12–16]	1972	126	1846
		T5	[16–20]	3768	245	3523
		T6	[20–24]	2766	416	2350

Tabel 3.3 *road accident attributes* menjelaskan pembagian waktu yang digunakan dalam penelitian “*A data mining framework to analyze road accident data*” dilakukan dengan perbedaan waktu 4 jam (*Interval 4 hours*). Pembagian waktu dengan kategori T1-T6 dengan durasi 4 jam ini digunakan sebagai acuan pembagian waktu dalam penelitian ini



Gambar 3.16 Road Accident Map Chart

Sumber : (Ramos, 2015)

Penelitian dengan judul “*Detection of road accident accumulation zones with a visual analytic approach*” digunakan sebagai pedoman bagaimana cara membuat dan melakukan pemetaan titik lokasi rawan kecelakaan dengan grafik *map*. Gambar 3.16 dijadikan pedoman dalam membuat grafik peta dalam penelitian ini. Penelitian dengan judul “*Analysis of road accidents in Poland over the last ten years*” digunakan sebagai pedoman dalam membuat data visual dan *dashboard*.

Perbedaan yang ada di dalam penelitian dengan ketiga penelitian tersebut adalah sumber data penelitian ini berasal dari *social media Twitter*, lokasi yang dijadikan objek penelitian adalah negara Indonesia di area Jabodetabek pada tahun 2014-2017. Penggunaan data dan variabel yang berbeda menghasilkan hasil grafik, *cluster*, dan analisa yang berbeda dari ketiga penelitian tersebut.

3.3 Perbandingan Metode

Tabel 3.4 Perbandingan Metode Visualisasi Data

<i>CRISP-DM</i>	<i>Data Mining</i>
-6 tahap yaitu <i>business understanding, data understanding, data preparation, Modelling, Evaluation, Deployment.</i>	-7 tahap yaitu <i>data cleaning, data intergration, selection data, transformation data, evaluation process, knowledge presentation.</i>
-Sumber data bisa tidak terstruktur	-Sumber data harus sudah terstruktur
-Sumber data bisa dari tipe penyimpanan apapun	-Data biasanya bersumber dari database

Tabel 3.4 perbandingan metode visualisasi data menjelaskan bahwa *CRISP-DM* memiliki kelebihan dalam sumber data bisa tidak terstruktur dan sumber data bisa dari tipe penyimpanan apa saja sehingga *CRISP-DM* lebih sesuai digunakan dalam penelitian ini dibanding tahap *data mining*. *Data mining* lebih sesuai digunakan untuk penelitian yang mengambil data dari *database*.

CRISP-DM memiliki satu tahap penting dalam penelitian ini yaitu tahap *data preparation* dimana persiapan data memiliki peranan untuk mendapatkan data dari *Twitter* karena data *Twitter* yang tidak terstruktur maka perlunya tahap persiapan data untuk mengolah data *tweets* menjadi informasi data informatif.

Metode yang dipilih adalah *CRISP-DM* sebagai metode tahap data visual karena tahapan yang ada pada *CRISP-DM* lebih sesuai untuk penelitian ini

dibanding metode *data mining*. *CRISP-DM* ini sendiri terdiri dari 6 tahap yaitu *business understanding*, *data understanding*, *data preparation*, *Modelling*, *Evaluation*, dan *Deployment*.

Tabel 3.5 Perbandingan Algoritma

<i>K-means</i>	<i>Naive Bayes</i>
- <i>Clustering algorithm</i>	- <i>Classification algorithm</i>
- <i>Unsupervised Learning</i>	- <i>Supervised Learning</i>
-Tipe data bisa angka dan teks	-Tipe Data lebih cenderung ke angka
-Hasil Analisa Pengelompokkan	-Hasil Analisa statistik dan prediksi

Tabel 3.5 perbandingan algoritma menjelaskan bahwa *k-means* merupakan algoritma klusterisasi sedangkan *naive bayes* merupakan algoritma klasifikasi. Klasifikasi merupakan proses menggabungkan beberapa objek menjadi objek baru yang lebih umum atau *general* sedangkan klusterisasi merupakan proses mengelompokkan objek sesuai hubungan atau kesamaan antar objek tersebut.

Perbedaan kedua yaitu *unsupervised learning* dan *supervised learning*. *K-means* sebagai algoritma *unsupervised learning* memiliki *output* yang tidak pasti tujuan dari *unsupervised learning* adalah untuk mendapatkan model dan struktur di dalam data untuk lebih mengenal data. *Naive bayes* merupakan kategori *supervised learning* dimana *supervised learning* memiliki *output* yang pasti dan sudah terstruktur. Algoritma *supervised learning* memiliki proses yang sudah pasti.

Metode *naive bayes* lebih sesuai digunakan untuk menganalisa angka dan probabilitas dimana data kecelakaan lalu lintas ini kurang sesuai dengan algoritma

naive bayes. Klasifikasi untuk data dalam penelitian ini tidak bisa dilakukan karena hasil dari klasifikasi menggunakan metode *naive bayes* tidak menghasilkan *output* apapun (penelitian menggunakan metode ini sudah dilakukan namun hasil klasifikasi tidak ada) karena data yang digunakan tidak sesuai dengan metode *naive bayes* dimana metode ini membutuhkan data yang bisa diklasifikasi dan terstruktur.

Metode yang digunakan untuk *Modelling* atau algoritma adalah *k-means*, metode dipilih karena tipe data dalam penelitian ini adalah teks. Penelitian ini memiliki variabel data yang bisa dikelompokkan secara *cluster* seperti wilayah, jenis kendaraan dan waktu kejadian yang bertujuan untuk mendapat pengelompokan *cluster* yang memiliki kesamaan antar data untuk menghasilkan analisa *cluster*.

UMMN

3.4 Perbandingan Tools

Microsoft Power BI FEATURES

- ✓ Datasets
- ✓ Customizable Dashboards
- ✓ Reports
- ✓ Navigation Pane
- ✓ Q&A Question Box
- ✓ Help and Feedback Buttons
- ✓ Ad Hoc Analysis
- ✓ Online Analytical Processing (OLAP)
- ✓ Trend Indicators
- ✓ Ad Hoc Reporting
- ✓ View rich graphical visualizations from complex BI data
- ✓ Full suite of BI tools for data visualization and reporting
- ✓ Author rich interactive reports with Power BI Desktop

Show Less

Pentaho FEATURES

- ✓ Data integration
- ✓ Business Analytics
- ✓ Big Data Analytics
- ✓ Embedded Analytics
- ✓ Cloud Analytics
- ✓ Ad Hoc Analysis
- ✓ Online Analytical Processing (OLAP)
- ✓ Predictive Analysis
- ✓ User-Friendly Interface
- ✓ Ad Hoc Reporting
- ✓ Customizable Features
- ✓ Performance Measurements
- ✓ Intuitive dashboards

Show Less

Gambar 3.17 Power BI Comparison

Sumber : (Finances Online, 2017)

Power BI dipilih sebagai *tools* dalam penelitian ini karena penggunaan *Power BI* sangat mudah dipelajari di banding *tools* lainnya. *Support* yang dimiliki *Power BI* lebih mudah di dapat karena *Power BI* memiliki forum diskusi aktif dan *training* yang dapat dipeleajari oleh *user* baru di website resmi *Power BI* dibanding *tools* lainnya.

Power BI juga memiliki keunikan yaitu fitur *custom visual*. *Custom visual* ini merupakan fitur untuk komunitas *developer Power BI* dari pengguna atau luar

maupun tim *microsoft* untuk mengeluarkan sebuah visual yang dikustom sesuai keinginan *developer* atau pembuat visual tersebut. *Custom visual* ini ada yang gratis hingga berbayar dan penelitian ini menggunakan beberapa fitur dari *custom visual* untuk lebih mempermudah visualisasi data dan menampilkan data dengan visual data yang lebih menarik dan mudah daripada *tools* lainnya.

3.5 Metode CRISP-DM

Penelitian ini menggunakan metode *CRISP-DM (Cross-Industry Standard Process for Data mining)* untuk melakukan proses *data mining* terhadap data kecelakaan lalu lintas di Jabodetabek. Tahap yang dilakukan terdiri dari 6 proses utama.

3.5.1 Business Understanding

Melakukan analisa objektif dan kebutuhan dari segi bisnis dan mengkonversi analisa tersebut menjadi cara memproses data dan membuat plan bagaimana mendapat objektif tersebut. Tujuan yang dimaksud dari penelitian ini adalah untuk mengurangi angka kecelakaan di Jabodetabek dan angka kematian akibat kecelakaan lalu lintas.

3.5.2 Data Understanding

Melakukan analisa data apa saja yang dibutuhkan untuk mengidentifikasi masalah yang perlu diselesaikan, menemukan solusi untuk mengolah data tersebut, membuat bagaimana cara membuat hipotesis data, dan dimana data tersebut dapat diambil.

3.5.3 *Data Preparation*

Data bersumber dari *social media Twitter* diproses dengan *text mining*, data yang diambil berasal dari *tweets* dengan kata kecelakaan. Kemudian data dilabel secara otomatis dengan menggunakan program *PHP* yang sudah dibuat.

3.5.4 *Modelling*

Data mentah (*raw data*) diolah dengan *modelling tools* yang dipilih. Pada tahap ini teknik *modelling* dipilih dan diterapkan. Data disesuaikan dengan kebutuhan teknik *modelling* maka terkadang *data preparation* dilakukan kembali untuk menyesuaikan teknik yang digunakan. Metode yang digunakan adalah *k-means*.

3.5.5 *Evaluation*

Data menghasilkan data visual dan *model* yang memiliki kualitas informasi yang baik sebelum melakukan *Deployment model* ini maka perlu dilakukan evaluasi *model* dan *review* step yang sudah dilakukan apakah sudah sesuai objektif apa belum oleh pihak lain khususnya penguji.

3.5.6 *Deployment*

Setelah data sudah diolah dan hasil yang dilakukan sudah informatif perlu adanya pembagian hasil tersebut ke orang luas dan orang yang menggunakannya. Pada tahap ini dilanjutkan dengan mempresentasikan informasi ke pengguna data tersebut dan penerapan tujuan informasi tersebut