

Analisis Big Data untuk Kota Aman

Setiyono
Husein Sastranegara Air Force Base
Indonesian Air Force
Bandung, Indonesia
Setiyono2717@gmail.com

Abstract—Smart solutions are needed by the city government to overcome various city problems. One solution is smart city. To realize smart city, one of the main challenges is the solution to overcome the city's security problems. Currently cities in Indonesia do not yet know the level of security of their cities. The level of city security can be obtained by surveying various cities. But surveys require personnel, time and cost that is not small. In this study the authors propose a method by designing a model to determine the level of security of cities in Indonesia by utilizing big data through the prediction of sentiment analysis of people's perceptions of city security on Twitter. This research was conducted in 25 cities in Indonesia which are divided into 8 big cities, 9 medium cities and 8 small cities. The results of the prediction models designed in this study are generally not much different from the results of the 2019 RKCI (Indonesia Smart Cities Rating) survey in the field of security and disaster. The results of this study found that 4 cities with a maturity level of security are at the Integrative level (score 60 to 79 in GSCM Maturity Level), namely Tangerang, Kediri, Parepare and Probolinggo, while the other 21 cities are at the Scattered level (score 40 to 59). The average score for the big city category is 55.41, while the middle city score is 55.48 and the small city is 53.70. The results of performance measurement of this prediction model are for an accuracy value of 80.10% while a precision value of 81.10% and a recall value of 82.62%.

Keywords—safe city, big data analytics, maturity level.

I. PENDAHULUAN

Pertumbuhan penduduk di dunia terus meningkat dan sekitar 68% dari populasi dunia pada tahun 2050 akan tinggal di daerah perkotaan. Di Indonesia sendiri saat ini tidak kurang dari 67% penduduk Indonesia tinggal di area perkotaan dan diramalkan pada tahun 2045 akan mencapai lebih dari 80%. Mau tidak mau hal ini akan memberikan dampak yang signifikan kepada kondisi kota. Tidak hanya dampak positif tetapi juga dampak negatif yang akan menimbulkan masalah-masalah perkotaan yang kompleks.

Solusi cerdas diperlukan dari pemerintah kota untuk mengatasi masalah ini. Salah satu solusi berlingkup luas yang bersifat cerdas dan inovatif yang ditawarkan saat ini adalah konsep *smart city* atau kota cerdas [1]. Menurut Lacinak [2] kota cerdas dengan integrasi teknologi dan lingkungan alami meningkatkan efektivitas proses di setiap bidang fungsinya, dalam rangka mencapai pembangunan berkelanjutan, keamanan dan kesehatan penduduk dengan tujuan untuk meningkatkan kualitas hidup warga, masyarakat dan lingkungan hidup. Pembangunan kota cerdas adalah proses, mentransformasikan kota menjadi kota cerdas melalui penerapan sistem konsep kota cerdas. Saat ini banyak kota berlomba-lomba menuju kota cerdas. Konsep kota cerdas adalah bagaimana mengelola kota dalam mengatasi masalah perkotaan yang disebabkan oleh populasi penduduk yang terus meningkat, seperti kurangnya kualitas layanan publik, kemacetan, masalah sampah, masalah ancaman keamanan maupun permasalahan kota lainnya [3]. Untuk mewujudkan kota cerdas salah satu tantangan utamanya adalah solusi untuk masalah keamanan kota.

Data-data tentang keamanan di kota-kota Indonesia saat ini belum terintegrasi dan dimanfaatkan secara optimal oleh pemerintah kota. Sumber data tersebut diantaranya dari data kepolisian, dinas-dinas pemerintah kota maupun data dari sumber lain seperti hasil survei maupun dari sumber internet antara lain situs berita maupun media sosial seperti facebook dan twitter.

Data-data tersebut sangat besar jumlahnya (*volume*), sangat beragam jenisnya (*variety*) dan sangat cepat pertumbuhannya (*velocity*). Lautan data ini mengarah pada satu terminologi *big data*. Ini juga dapat merujuk pada proses menggunakan analisis prediktif, analisis perilaku pengguna atau teknologi analisis data lanjutan lainnya untuk mengekstrak nilai dari kumpulan data. Big data sering digunakan di bisnis atau instansi pemerintah untuk menemukan tren dan pola, yang dapat membantu mereka mengambil keputusan strategis atau menemukan pola atau tren tertentu.

Beberapa manfaat *big data* yang sudah dirasakan khususnya bagi dunia usaha diantaranya untuk mengetahui respons masyarakat terhadap produk-produk yang dikeluarkan melalui analisis sentimen di media sosial, membantu perusahaan mengambil keputusan secara lebih tepat dan akurat berdasarkan data, membantu meningkatkan citra perusahaan di mata pelanggan, untuk perencanaan usaha dengan mengetahui perilaku pelanggan, seperti pada perusahaan telekomunikasi dan perbankan, serta mengetahui tren pasar dan keinginan konsumen. Selain bermanfaat untuk analisis bisnis, teknologi *big data* juga dapat dimanfaatkan secara luas di pemerintahan. Beberapa peluang pemanfaatan big data di sektor publik antara lain untuk mendapatkan feedback dan respon masyarakat dari sistem informasi layanan pemerintah maupun dari media sosial, sebagai dasar penyusunan kebijakan dan perbaikan pelayanan publik, menemukan solusi atas permasalahan yang ada berdasarkan data.

Penggunaan media sosial yang signifikan saat ini menghasilkan data yang sangat besar yang tersedia secara online. Twitter merupakan media sosial yang sering digunakan datanya untuk memprediksi hasil pemilihan presiden atau kepala daerah. Menurut Soler [4][5], penggunaan media sosial twitter sebagai alat utama prediksi terbukti sebagai sumber yang valid dan efektif. Prediksi hasil pemilihan di beberapa negara dengan menggunakan data twitter lebih efektif baik dalam biaya maupun waktu dibandingkan dengan cara polling offline konvensional.

Kumpulan data *tweet* dapat dianalisis untuk mendapatkan informasi yang berguna dengan menggunakan analisis sentimen. Analisis sentimen atau disebut juga *opinion mining* merupakan salah satu bagian *text mining*. *Text mining* sendiri merupakan bagian dari *data mining* yang proses utamanya adalah melakukan ekstraksi pengetahuan dan informasi dari pola yang terdapat pada kumpulan dokumen teks menggunakan alat analisis tertentu [6]. Analisis sentimen dilaksanakan dengan melakukan studi mengenai opini orang-orang, evaluasi, sentimen, tingkah laku serta emosi pada suatu entitas seperti produk, layanan, individu, organisasi, permasalahan, topik, acara dan atribut-atributnya [7]. Analisis sentimen juga dikenal sebagai deteksi subjektivitas atau penambangan opini untuk menemukan pola linguistik dan statistik menggunakan metode pembelajaran mesin dan pemrosesan bahasa alami [8].

Pemanfaatan *big data analytics* contohnya analisis sentimen dari media sosial seperti twitter cukup banyak digunakan di dunia bisnis. Analisis Sentimen adalah metode untuk menilai sentimen atau perasaan seseorang sehubungan dengan hal tertentu. Ini digunakan untuk mengenali dan mengatur sentimen yang dikomunikasikan dalam tulisan. Pang [9] melakukan penelitian analisis sentimen dengan menggunakan data *movie review* milik Cornell University. Penelitian ini mengklasifikasikan review film sebagai sentimen positif atau sentimen negatif. Untuk menentukan polaritasnya, peneliti mendefinisikan strategi pembelajaran mesin yang menerapkan strategi klasifikasi konten hanya pada bagian subjektif dari dokumen teks. Metode yang digunakan adalah sebuah review film dikatakan positif ketika pemilik review memberi rating > 50% dari rating maksimum. Apabila rating \leq 50% dari rating maksimum, maka review tersebut bernilai negatif.

Rumusan masalah dalam penelitian ini adalah bagaimana merancang model untuk mengetahui tingkat keamanan kota di Indonesia dengan memanfaatkan *big data* melalui prediksi analisis sentimen persepsi masyarakat di twitter dan bagaimana melakukan evaluasi rancangan dan kinerja model tersebut serta memvalidasi hasilnya. Tujuan dari penelitian ini adalah merancang model untuk mengetahui tingkat keamanan kota di Indonesia dengan memanfaatkan *big data*.

Dalam melakukan penelitian ini penulis membatasi masalah penelitian sebagai berikut.

- 1) Penelitian ini menggunakan studi kasus keamanan kota pada 25 kota di Indonesia.
- 2) *Score* tingkat keamanan kota diperoleh dengan menghitung jumlah persentase sentimen positif dari persepsi masyarakat di twitter tentang keamanan kota.

- 3) *Tweet* yang digunakan adalah *tweet* berbahasa Indonesia.
- 4) *Tweet* yang dianalisis menggunakan beberapa kata kunci tertentu yang berhubungan dengan keamanan kota.
- 5) *Tweet* diambil selama 1 tahun yaitu dari 1 Januari 2018 sampai dengan 31 Desember 2018.
- 6) Algoritma klasifikasi yang digunakan adalah *logistic regression*.

II. METODE PENELITIAN

A. Algoritma Klasifikasi *Logistic Regression*

Analisis sentimen pada penelitian ini dilakukan dengan menggunakan metode algoritma klasifikasi pembelajaran mesin *logistic regression*. *Logistic Regression* adalah salah satu algoritma pembelajaran mesin yang paling populer untuk klasifikasi biner. *Logistic Regression* digunakan ketika variabel *dependen* (target) adalah kategorikal. Beberapa contoh *logistic regression* adalah :

- 1) Untuk memprediksi apakah email itu spam (1) atau bukan spam (0)
- 2) Untuk mendeteksi penipuan, apakah transaksi curang (1) atau tidak curang (0)

Algoritma klasifikasi *logistic regression* digunakan untuk menetapkan pengamatan pada sekumpulan kelas diskrit. Tidak seperti regresi linier yang menghasilkan nilai angka kontinyu, *logistic regression* mengubah outputnya menggunakan fungsi sigmoid logistik untuk mengembalikan nilai probabilitas yang kemudian dapat dipetakan ke dua atau lebih kelas diskrit. Dalam regresi logistik di kenal adanya fungsi logistik. Fungsi logistik juga disebut fungsi sigmoid dikembangkan oleh ahli statistik untuk menggambarkan sifat-sifat pertumbuhan populasi dalam ekologi, meningkat dengan cepat dan memaksimalkan kemampuan daya dukung lingkungan. Ini adalah kurva berbentuk S yang dapat mengambil angka bernilai nyata dan memetakannya menjadi nilai antara 0 dan 1, tetapi tidak pernah persis pada batas itu. Formula dari fungsi logistik adalah seperti pada persamaan 1, sebagai berikut:

$$\text{Sig}(t) = 1 / (1 + e^{-t}) \quad (1)$$

Di mana e adalah basis dari logaritma natural (angka Euler atau fungsi EXP ()) dan nilainya adalah nilai numerik aktual yang ingin diubah. Untuk memprediksi kelas, ambang batas dapat ditetapkan. Berdasarkan ambang ini, estimasi probabilitas yang diperoleh diklasifikasikan ke dalam kelas-kelas. Katakanlah, jika $\text{predict_value} \geq 0,5$, maka klasifikasikan email sebagai spam, sedangkan jika $\text{predict_value} < 0,5$ sebagai bukan spam. Algoritma klasifikasi *logistic regression* ini dipilih oleh penulis juga karena berdasar penelitian sebelumnya hasil akurasi prediksinya baik.

B. Evaluasi Model

Untuk memvalidasi algoritma prediksi sentimen pada penelitian ini digunakan metode *k fold cross validation* sehingga bisa diketahui akurasi prediksi dari model yang telah dibuat. *Cross validation* (validasi silang) adalah metode statistik yang digunakan untuk memperkirakan keterampilan model pembelajaran mesin. Metode ini biasanya digunakan dalam pembelajaran mesin yang diterapkan untuk membandingkan dan memilih model untuk masalah pemodelan prediktif yang diberikan karena mudah dipahami dan diimplementasikan, serta menghasilkan estimasi keterampilan yang umumnya memiliki bias yang lebih rendah daripada metode lainnya.

K fold cross validation adalah tipe umum validasi silang yang banyak digunakan dalam pembelajaran mesin. Biasanya peneliti menggunakan $k=5$ atau $k=10$. Dalam penelitian ini digunakan $k=10$ karena biasanya lebih rendah dibanding $k=5$. Penulis dalam penelitian ini menggunakan metode *10 fold cross validation*. Dalam *10 fold cross validation*, dataset akan dibagi menjadi 10 bagian, yaitu 9 bagian digunakan untuk proses pelatihan/*training* dan 1 bagian lainnya digunakan untuk proses pengujian/*testing* [10]. Iterasi dilakukan sebanyak 10 kali dengan menggunakan kombinasi 10 bagian data tersebut.

Untuk mengukur kinerja dari algoritma klasifikasi, penulis menggunakan metode *confusion matrix*. Dalam *confusion matrix*, hasil klasifikasi oleh sistem dibandingkan dengan hasil klasifikasi seharusnya [11]. Pada *confusion matrix*, ada 4 istilah yang merupakan representasi hasil dari proses klasifikasi. Istilah-istilah tersebut yaitu sebagai berikut :

- 1) True Positive (TP), merupakan data aktual positif yang diprediksi positif
- 2) True Negative (TN), merupakan data aktual negatif yang diprediksii negatif
- 3) False Positive (FP), merupakan data aktual negatif namun diprediksi positif
- 4) False Negative (FN), merupakan data aktual positif namun diprediksi negatif

Dari nilai TP, TN, FP dan FN dapat dihasilkan nilai dari akurasi, presisi serta *recall*. Akurasi adalah perbandingan antara data terklasifikasi secara benar dengan keseluruhan dari data (persamaan 2). Sedangkan presisi merupakan jumlah data positif yang diklasifikasikan benar dibagi total data yang diklasifikasikan positif (persamaan 3). *Recall* menggambarkan berapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem (persamaan 4). Untuk persamaan dari akurasi, presisi dan recall ditunjukkan seperti persamaan di bawah ini.

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100 \% \quad (2)$$

$$\text{Presisi} = \frac{\text{TP}}{\text{FP} + \text{TP}} \times 100 \% \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}} \times 100 \% \quad (4)$$

Keterangan : TP = True Positif FP = False Positif
 TN = True Negatif FN = False Negatif

C. Pengukuran Tingkat Kematangan Keamanan Kota

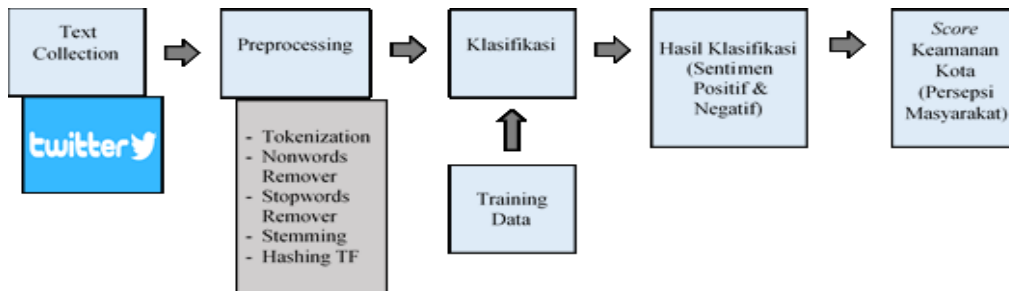
Pengukuran tingkat kematangan keamanan kota pada penelitian ini menggunakan *Garuda Smart City Model* (GSCM). Dalam GSCM tingkat kematangan keamanan kota terdapat 5 level kematangan [3], yaitu:

- 1) Level *Adhoc* (*score* <= 20), di mana pemerintah kota memiliki keinginan membuat kotanya menjadi aman, tetapi belum merencanakan dan mengupayakan.
- 2) Level *Initiative* (*score* 21-39), di mana pemerintah kota memiliki keinginan dan mulai melakukan perencanaan untuk konsep kota aman.
- 3) Level *Scattered* (*score* 40-59), yaitu konsep pembangunan kota aman mulai berjalan, tetapi belum terintegrasi.
- 4) Level *Integrative* (*score* 60-79), yaitu ketika pelaksanaan konsep kota aman telah terintegrasi.
- 5) Level *Smart* (*score* >= 80), yaitu saat perencanaan dan pembangunan kota aman berlangsung secara terintegrasi dan berkelanjutan.

III. ANALISIS

Analisis sentimen pada penelitian ini menggunakan Apache Spark 2.2.0 yang dilakukan dengan merancang model untuk mengklasifikasikan suatu *tweet* apakah isi *tweet* tersebut termasuk sentimen positif atau sentimen negatif terhadap situasi keamanan suatu kota, sehingga kita dapat mengetahui bagaimana persepsi masyarakat suatu kota terhadap tingkat keamanan

kotanya. Model prediksi analisis sentimen persepsi masyarakat tentang keamanan kota seperti ditunjukkan pada gambar 1.



Gambar 1. Model prediksi analisis sentimen persepsi masyarakat tentang keamanan kota

Proses analisis data mengenai persepsi masyarakat pada penelitian ini dilakukan dengan 4 proses pemberdayaan *big data analytics* yaitu pengumpulan data (*data collecting*), penyimpanan data (*data storing*), analisis data (*data analytics*) dan visualisasi (*data visualisation*) [12].

A. Pengumpulan Data (*data collecting*)

Sumber data yang digunakan adalah *tweet* dari twitter tentang keamanan 25 kota di indonesia. Kota-kota tersebut adalah 8 kota besar (Surabaya, Bekasi, Bandung, Semarang, Tangerang Selatan, Batam, Bogor, Tangerang), 9 kota sedang (Denpasar, Yogyakarta, Kediri, Surakarta, Jambi, Madiun, Pontianak, Probolinggo, Samarinda) dan 8 kota kecil (Bontang, Magelang, Tual, Mojokerto, Padangpanjang, Pariaman, Parepare, Salatiga). *Tweet* yang diambil adalah *tweet* yang mengandung kata kunci (*keyword*) yang relevan dengan keamanan kota dan sesuai dengan indikator keamanan pada *Garuda Smart City Model* (GSCM) serta ISO 37120 (*Sustainable development of communities: Indicators for city services and quality of life*). *Keyword* tersebut yaitu “(nama kota) aman”, “(nama kota) kondusif”, “(nama kota) terorisme”, “(nama kota) banjir”, “(nama kota) bencana”, “(nama kota) gempa”, “(nama kota) pengemis”, “(nama kota) gelandangan”, “(nama kota) kebakaran”, “(nama kota) pencurian”, “(nama kota) kerusuhan”, “(nama kota) siskamling. Struktur data *tweet* yang diambil adalah sebagai berikut *username, tweet text, mentions, retweets, favourites, geo, hashtags* dan *permalink*.

Data-data twitter ini dapat diperoleh dari twitter API, tetapi twitter API versi gratis hanya memberikan akses unduhan *tweet* maksimal satu minggu ke belakang, sementara penelitian ini mengambil data *tweet* selama satu tahun dari 1 januari 2018 sampai 31 Desember 2018. Hal ini dapat diatasi menggunakan tool buatan Jefferson Henrique berbasis python yang dapat diunduh dari github untuk *crawling* data *tweet*. Hasil *crawling* *tweet* ini dalam format CSV. Dari hasil *crawling* ini didapatkan total 98.263 *tweet*. Kemudian diambil 1.100 *tweet* untuk pembuatan model klasifikasi, dan dilabeli sentimennya secara manual. Setelah dilakukan pembersihan dengan menghilangkan *tweet* duplikat dan *tweet* dari bot diperoleh total *tweet* 74.717. Rincian data jumlah *tweet* 74.717 seperti ditunjukkan pada table 1, di mana *tweet* terbanyak adalah Surabaya dan *tweet* paling sedikit adalah Padangpanjang.

Tabel 1.

Data jumlah *tweet* tentang keamanan dari hasil *crawling* 25 kota di Indonesia

Kategori Kota	Nama Kota	Jumlah <i>Tweet</i>
Kota Besar	Surabaya	12.729
	Bekasi	5.324
	Bandung	9.774
	Semarang	9.039
	Tangerang Selatan	1.479

	Batam	1.449
	Bogor	9.689
	Tangerang	4.655
Kota Sedang	Denpasar	3.178
	Jambi	3.100
	Yogyakarta	4.017
	Kediri	1.836
	Surakarta	301
	Madiun	1.037
	Pontianak	1.141
	Probolinggo	1.554
	Samarinda	1.234
Kota Kecil	Bontang	821
	Magelang	861
	Mojokerto	672
	Tual	122
	Padangpanjang	80
	Pariaman	395
	Parepare	230
	Salatiga	229
	Jumlah <i>tweet</i>	74.717

B. Penyimpanan data (*data storing*)

Data *tweet* hasil *crawling* di simpan dalam database SQLspark. Data 25 kota dipisahkan sehingga ada 25 file database.

C. Analisis data (*data analytics*)

Langkah-langkah dari proses untuk mendapatkan hasil analisis sentimen adalah sebagai berikut :

1) Meyiapkan training data

Training data ini akan digunakan dalam proses pembuatan model untuk memprediksi apakah isi suatu *tweet* mengandung kategori sentimen positif atau sentimen negatif. Training data ini merupakan kumpulan teks *tweet* yang telah diberi label sentimen (positif atau negatif) secara manual. Pada penelitian ini jumlah data trainingnya adalah 1100 *tweet*.

2) Memproses training data.

Training data diproses untuk mendapatkan *feature words*. Prosesnya yaitu teks *tweet* dalam training data sebelumnya dibersihkan (*cleaning*) dulu dan disaring (*filtering*) untuk menghilangkan noise (unsur yang tidak diperlukan). *Pre-processing* data ini adalah salah satu langkah penting saat melakukan analisis sentimen. Tujuan utama dari *pre-processing* adalah untuk meningkatkan kualitas data dengan menghapus kata-kata yang tidak diinginkan, keuntungannya adalah ada pengurangan dalam ruang ukuran fitur [13]. Metode *pre-processing* yang digunakan dalam penelitian ini adalah sebagai berikut:

- a) *Teks tokenization* dan *case folding*, yaitu memecah kalimat *tweet* menjadi token (unsur tunggal berdiri sendiri) dan mengubah huruf besar menjadi huruf kecil.

Tabel 2.
Tokenization dan Case Folding

Teks <i>Tweet</i> (Sebelum <i>Preprocessing</i>)	Teks <i>Tweet</i> Setelah <i>Preprocessing</i> (<i>Tokenization</i> dan <i>Case folding</i>)
Bandung dari siang sampai ini malam aman . Dan terkesan sangat sepi .	[bandung, dari, siang, sampai, ini, malam, aman, ., dan, terkesan, sangat, sepi, .]
# bandungbanjir hati hati yang arah ke Sukajadi . kendaraan kecil sebaiknya menghindar .	[#, bandungbanjir, hati, hati, yang, arah, ke, sukajadi, ., kendaraan, kecil, sebaiknya, menghindar, .]

b) Pembersihan *nonwords characters*, yaitu seperti tanda baca, tagar.

Tabel 3.
Nonwords Remover

Teks <i>Tweet</i> Hasil <i>Tokenization</i> dan <i>Case Folding</i>	Teks <i>Tweet</i> Setelah <i>Preprocessing</i> (<i>Nonwords Remover</i>)
[bandung, dari, siang, sampai, ini, malam, aman, ., dan, terkesan, sangat, sepi, .]	[bandung, dari, siang, sampai, ini, malam, aman, dan, terkesan, sangat, sepi]
[#, bandungbanjir, hati, hati, yang, arah, ke, sukajadi, ., kendaraan, kecil, sebaiknya, menghindar, .]	[bandungbanjir, hati, hati, yang, arah, ke, sukajadi, kendaraan, kecil, sebaiknya, menghindar]

c) Pembersihan *stopwords*, yaitu menghapus kata-kata sambung seperti dan, atau, dari, ke, daripada dan lain-lain.

Tabel 4.
Stopwords Remover

Teks <i>Tweet</i> Hasil <i>Nonwords Remover</i>	Teks <i>Tweet</i> Setelah <i>Preprocessing</i> (<i>Stopwords Remover</i>)
[bandung, dari, siang, sampai, ini, malam, aman, dan, terkesan, sangat, sepi]	[bandung, siang, malam, aman, terkesan, sepi]
[bandungbanjir, hati, hati, yang, arah, ke, sukajadi, kendaraan, kecil, sebaiknya, menghindar]	[bandungbanjir, hati, hati, arah, sukajadi, kendaraan, menghindar]

d) *Stemming*, merupakan proses menemukan kata dasar dari suatu kata, yaitu dengan menghilangkan imbuhan, baik yang terdiri dari akhiran (*suffixes*), sisipan (*infixes*) maupun awalan (*prefixes*), ataupun kombinasi dari awalan serta akhiran (*confixes*) pada suatu kata turunan. Penggunaan *stemming* adalah untuk mengganti bentuk dari sebuah kata menjadi kata dasarnya. Hasil dari proses *stemming* ini merupakan *feature words*.

Tabel 5.
Stemming

Teks <i>Tweet</i> Hasil <i>Stopwords Remover</i>	Teks <i>Tweet</i> Setelah <i>Preprocessing</i> (<i>Stemming</i>)
[bandung, siang, malam, aman, terkesan, sepi]	[bandung, siang, malam, aman, kesan, sepi]
[bandungbanjir, hati, hati, arah, sukajadi, kendaraan, menghindar]	[bandungbanjir, hati, hati, arah, sukajadi, kendaraan, hindar]

e) *Feature words* dan *hashing TF*, di mana *feature words* akan menjadi fitur dalam melakukan analisis sentimen. Dalam *machine learning* data teks akan diubah menjadi numerik. Apache spark telah mengimplementasikan algoritma untuk mengubah data teks menjadi data numerik dalam sebuah fungsi yaitu *hashing TF*.

Tabel 6.
Hashing TF

Teks Tweet Hasil Stemming	Teks Tweet Setelah Preprocessing (Hashing TF)
[bandung, siang, malam, aman, kesan, sepi]	(262144,[59034,41100,138013,164586,174549,238366],[1.0,1.0,1.0,1.0,1.0,1.0])
[bandungbanjir, hati, hati, arah, sukajadi, kendaraan, hindar]	(262144,[10123,15772,147296,178673,233526,234821],[2.0,1.0,1.0,1.0,1.0,1.0])

- 3) Penerapan algoritma klasifikasi *logistic regression*
Penerapan algoritma ini untuk menghasilkan model yang dapat diaplikasikan pada teks *tweet* mentah untuk menentukan sentimen yang terkandung dalam *tweet* tersebut, apakah sentimen positif atau sentimen negatif. Dalam tahap ini, algoritma klasifikasi *logistic regression* akan melalui proses training, yaitu proses mempelajari karakteristik *feature words* dari training data yang nantinya akan menjadi model untuk dijadikan acuan pada proses klasifikasi.
- 4) Evaluasi model
Evaluasi model adalah dengan validasi dan mengukur kinerja dari algoritma model yang digunakan berdasarkan hasil prediksi sentimen yang diperoleh. Pengukuran ini didapatkan dengan melakukan pengujian menggunakan data testing. Pada penelitian ini validasi menggunakan metode *10 fold cross validation*, sedangkan pengukuran kinerja algoritma klasifikasi menggunakan *confusion matrix*.

D. Visualisasi (data visualisation)

Visualisasi dilakukan agar data-data yang dihasilkan dari proses analisa dapat dimengerti dengan mudah, misalnya dalam bentuk *chart*.

IV. HASIL

A. Lima Besar Keyword Hasil Crawling Tweet Tiap Kota

Pada tabel 7 di tunjukkan 5 besar jumlah *tweet* berdasar *keyword* hasil *crawling tweet* di 25 kota Indonesia. Tabel 7 ini menggambarkan *keyword* yang paling sering diungkapkan oleh masyarakat tentang kondisi keamanan suatu kota. Dari tabel ini tiap kota dapat mengetahui dan memetakan masalah-masalah keamanan kotanya yang sering diperbincangkan oleh masyarakat sehingga dapat dijadikan salah satu acuan untuk menentukan prioritas kebijakan bidang keamanan kotanya. Sebagai contoh kota Semarang, dari tabel 7 dapat dilihat bahwa *keyword* “semarang banjir” menduduki peringkat pertama percakapan *tweet* yang porsinya sampai 48,28% dari total keseluruhan *tweet*, ini berarti masalah banjir di kota semarang perlu menjadi perhatian serius dari Pemkot Semarang untuk ditangani.

Tabel 7.
Lima Besar Jumlah Tweet Berdasar Keyword di 25 kota Indonesia

No	Kategori Kota	Nama Kota	5 Besar Jumlah Tweet Berdasar Keyword
1	Kota Besar	Surabaya	1. “surabaya terorisme” (35,74%) 2. “surabaya gempa” (14,87%) 3. “surabaya aman” (12,55%) 4. “surabaya banjir” (8,25%) 5. “surabaya bencana” (6,35%)
2		Bekasi	1. “bekasi aman” (31,28%) 2. “bekasi banjir” (24,64%) 3. “bekasi gempa” (17,25%) 4. “bekasi kebakaran” (9,10%)

			5. "bekasi bencana" (6,06%)
3		Bandung	1. "bandung aman" (45,29%) 2. "bandung banjir" (27,78%) 3. "bandung kebakaran" (13,59%) 4. "bandung kondusif" (5,77%) 5. "bandung kerusuhan" (2,65%)
4		Semarang	1. "semarang banjir" (48,28%) 2. "semarang aman" (28,12%) 3. "semarang bencana" (7,95%) 4. "semarang gempa" (4,91%) 5. "semarang kebakaran" (4,89%)
5		Tangsel	1. "tangsel gempa" (27,32%) 2. "tangsel aman" (24,07%) 3. "tangsel kebakaran" (14,74%) 4. "tangsel banjir" (14,40%) 5. "tangsel pencurian" (9,74%)
6		Batam	1. "batam aman" (29,58%) 2. "batam kebakaran" (20,48%) 3. "batam banjir" (18,08%) 4. "batam gempa" (13,65%) 5. "batam bencana" (6,89%)
7		Bogor	1. "bogor banjir" (27,10%) 2. "bogor gempa" (26,70%) 3. "bogor bencana" (17,65%) 4. "bogor aman" (13,44%) 5. "bogor kebakaran" (8,72%)
8		Tangerang	1. "tangerang gempa" (23,42%) 2. "tangerang aman" (19,68%) 3. "tangerang banjir" (18,00%) 4. "tangerang kebakaran" (13,90%) 5. "tangerang bencana" (10,42%)
9	Kota Sedang	Denpasar	1. "denpasar gempa" (56,23%) 2. "denpasar aman" (12,62%) 3. "denpasar banjir" (11,77%) 4. "denpasar kebakaran" (7,74%) 5. "denpasar bencana" (6,48%)
10		Jambi	1. "jambi aman" (42,96%) 2. "jambi banjir" (23,02%) 3. "jambi gempa" (11,61%) 4. "jambi bencana" (5,66%) 5. "jambi kondusif" (4,55%)
11		Yogyakarta	1. "yogyakarta gempa" (38,98%) 2. "yogyakarta aman" (24,07%) 3. "yogyakarta bencana" (20,86%) 4. "yogyakarta kebakaran" (4,46%) 5. "yogyakarta banjir." (3,73%)
12		Kediri	1. "kediri bencana" (21,24%) 2. "kediri kebakaran" (14,32%) 3. "kediri aman" (13,83%) 4. "kediri kondusif" (12,69%) 5. "kediri banjir" (12,31%)
13		Surakarta	1. "surakarta bencana" (23,59%) 2. "surakarta kebakaran" (22,26%) 3. "surakarta aman" (19,27%) 4. "surakarta banjir" (13,62%) 5. "surakarta gempa" (7,97%)
14		Madiun	1. "madiun banjir" (28,74%) 2. "madiun kebakaran" (23,72%) 3. "madiun terorisme" (13,11%) 4. "madiun bencana" (10,61%) 5. "madiun pencurian" (8,29%)
15		Pontianak	1. "pontianak kebakaran" (35,50%) 2. "pontianak aman" (27,87%)

			3. "pontianak gempa" (11,39%) 4. "pontianak banjir" (10,08%) 5. "pontianak bencana" (8,41%)
16		Probolinggo	1. "probolinggo banjir" (23,62%) 2. "probolinggo bencana" (21,36%) 3. "probolinggo aman" (15,96%) 4. "probolinggo kebakaran" (11,07%) 5. "probolinggo pencurian" (10,36%)
17		Samarinda	1. "samarinda aman" (31,69%) 2. "samarinda kebakaran" (16,77%) 3. "samarinda gempa" (13,53%) 4. "samarinda pencurian" (12,97%) 5. "samarinda bencana" (9,24%)
18	Kota Kecil	Bontang	1. "bontang banjir" (36,18%) 2. "bontang gempa" (20,10%) 3. "bontang bencana" (15,10%) 4. "bontang kebakaran" (13,52%) 5. "bontang aman" (8,04%)
19		Magelang	1. "magelang bencana" (28,80%) 2. "magelang kebakaran" (23,00%) 3. "magelang gempa" (12,77%) 4. "magelang aman" (12,66%) 5. "magelang banjir" (7,43%)
20		Mojokerto	1. "mojokerto kebakaran" (42,26%) 2. "mojokerto aman" (19,35%) 3. "mojokerto banjir" (14,58%) 4. "mojokerto gempa" (11,31%) 5. "mojokerto pencurian" (7,14%)
21		Tual	1. "tual gempa" (42,62%) 2. "tual aman" (38,52%) 3. "tual konduktivitas" (7,38%) 4. "tual bencana" (7,38%) 5. "tual banjir" (1,64%)
22		Padang panjang	1. "padangpanjang gempa" (32,50%) 2. "padangpanjang bencana" (26,25%) 3. "padangpanjang aman" (20,00%) 4. "padangpanjang kebakaran" (10,00%) 5. "padangpanjang banjir" (7,50%)
23		Pariaman	1. "pariaman gempa" (27,09%) 2. "pariaman banjir" (18,48%) 3. "pariaman bencana" (17,22%) 4. "pariaman aman" (14,43%) 5. "pariaman kebakaran" (5,82%)
24		Parepare	1. "parepare kebakaran" (39,39%) 2. "parepare bencana" (13,64%) 3. "parepare pencurian" (13,13%) 4. "parepare gempa" (13,13%) 5. "parepare aman" (12,63%)
25		Salatiga	1. "salatiga kebakaran" (27,95%) 2. "salatiga aman" (21,83%) 3. "salatiga bencana" (13,54%) 4. "salatiga gempa" (12,66%) 5. "salatiga banjir" (8,30%)

B. Persentase Sentimen Positif dan Negatif Per Bulan Tiap Kota

Pada tabel 8, tabel 9 dan tabel 10 di tunjukkan jumlah sentimen positif dan sentimen negatif dari tiap kota untuk kategori kota besar, kota sedang dan kota kecil. Di tabel tersebut bisa dilihat dinamika perubahan sentimen tiap bulannya dari tiap kota, dari data ini juga dapat dijadikan bahan pemetaan masalah dan bahan evaluasi tiap kota untuk melihat bagaimana perkembangan persepsi publik terhadap keamanan kotanya. Sebagai contoh kota Semarang seperti digambarkan pada gambar 2 berdasar data tabel 8 terlihat bahwa pada bulan januari, february dan desember sentimen

negatif begitu tinggi dikarenakan di bulan-bulan tersebut adalah bulan dengan curah hujan cukup tinggi yang mengakibatkan banjir.

Tabel 8. Persentase Sentimen Positif dan Negatif di Twitter 8 Kota Besar Indonesia tiap Bulan pada Tahun 2018

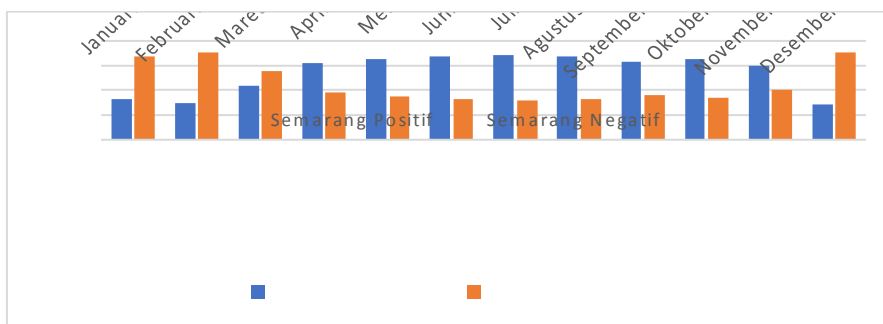
	KOTA BESAR															
	Surabaya		Bekasi		Bandung		Semarang		Tangsel		Batam		Bogor		Tangerang	
	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif
Januari	44,05	55,95	25,99	74,61	64,37	35,63	32,55	67,45	39,00	61,00	43,66	56,34	41,05	58,95	63,57	36,43
Februari	45,98	54,02	73,33	26,67	73,83	26,17	29,68	70,32	52,63	47,37	40,28	59,72	49,26	50,74	68,21	31,79
Maret	47,70	52,30	38,48	61,52	71,38	28,62	43,95	56,05	68,63	31,37	66,14	33,86	51,03	48,97	49,79	50,21
April	52,54	47,46	40,13	59,87	33,43	66,57	61,97	38,03	46,55	53,45	56,86	43,14	35,50	64,50	64,83	35,17
Mei	36,06	63,94	54,22	45,78	65,88	34,12	64,79	35,21	61,02	38,98	64,95	35,05	53,07	46,93	72,76	27,24
Juni	43,34	56,66	55,25	44,75	71,75	28,25	79,93	20,07	76,79	23,21	45,07	54,93	43,06	56,94	76,98	23,02
Juli	57,59	42,41	49,59	50,41	39,94	60,06	68,01	31,99	63,75	36,25	57,30	42,70	41,14	58,86	63,74	36,26
Agustus	61,89	38,11	55,74	44,26	63,11	36,89	66,99	33,01	60,14	39,86	53,27	46,73	60,86	39,14	68,47	31,53
September	60,59	39,41	42,82	57,18	54,33	45,67	63,58	36,42	71,10	28,90	63,50	36,50	57,66	42,34	76,37	23,63
Oktober	56,44	43,56	53,63	46,37	51,69	48,31	65,76	34,24	71,07	28,93	68,80	31,20	70,89	29,11	85,19	14,81
November	48,09	51,91	42,74	57,26	48,61	51,39	59,78	40,22	54,08	45,92	64,80	35,20	73,56	26,44	59,69	40,31
Desember	47,95	52,05	40,29	59,71	63,20	36,80	29,02	70,98	66,67	33,33	60,78	39,22	79,26	20,74	73,94	26,06

Tabel 9. Persentase Sentimen Positif dan Negatif di Twitter 9 Kota Sedang Indonesia tiap Bulan pada Tahun 2018

	KOTA SEDANG																	
	Denpasar		Jambi		Yogyakarta		Kediri		Surakarta		Madiun		Pontianak		Probolinggo		Samarinda	
	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif
Januari	38,64	61,36	30,30	69,70	38,85	61,15	53,97	46,03	40,00	60,00	55,56	44,44	41,93	58,07	92,65	7,35	81,97	18,03
Februari	61,18	38,82	26,97	73,03	47,13	52,87	73,44	26,56	30,00	70,00	45,56	54,44	54,57	45,43	37,70	62,30	50,00	50,00
Maret	73,26	26,74	16,84	83,16	60,22	39,78	47,37	52,63	71,43	28,57	27,12	72,88	56,52	43,48	75,00	25,00	65,57	34,43
April	78,81	21,19	43,18	56,82	62,21	37,79	47,66	52,34	52,94	47,06	89,47	10,53	45,33	54,67	81,25	18,75	60,66	39,34
Mei	70,34	29,66	66,47	33,53	65,45	34,55	70,34	29,66	70,00	30,00	59,09	40,91	53,85	46,15	71,43	28,57	73,91	26,09
Juni	71,77	28,23	84,65	15,35	61,22	38,78	67,01	32,91	30,00	70,00	61,90	38,10	46,94	53,06	91,86	8,14	55,42	44,58
Juli	54,89	45,11	57,95	42,05	67,71	32,29	38,38	61,62	57,14	42,86	65,85	34,15	55,68	44,32	48,21	51,79	55,88	44,12
Agustus	43,05	56,95	51,61	48,39	48,90	51,10	63,51	36,49	33,33	66,67	65,35	34,65	36,29	63,71	66,15	33,85	64,62	35,38
September	58,93	41,07	75,64	24,36	66,37	33,63	68,47	31,53	69,77	30,23	52,63	47,37	38,93	61,07	53,68	46,32	43,87	56,12
Oktober	47,57	52,43	63,99	36,01	63,50	36,50	85,01	14,99	46,97	53,03	67,29	32,71	55,79	44,21	60,38	39,62	51,40	48,60
November	44,78	55,22	45,08	54,92	36,91	63,09	76,97	23,03	39,53	60,47	78,95	21,05	56,42	43,58	78,85	21,15	60,29	39,71
Desember	27,56	72,44	41,42	58,58	54,64	45,36	71,97	28,03	58,33	41,67	57,61	42,39	44,08	55,92	47,33	52,67	22,33	77,67

Tabel 10. Persentase Sentimen Positif dan Negatif di Twitter 8 Kota Kecil Indonesia tiap Bulan pada Tahun 2018

	KOTA KECIL															
	Bontang		Magelang		Mojokerto		Tual		Padangpanjang		Pariaman		Parepare		Salatiga	
	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif	Positif	Negatif
Januari	32,86	67,14	55,77	44,23	41,67	58,33	44,44	55,56	62,50	37,50	81,82	18,18	53,85	46,15	33,33	66,67
Februari	66,67	33,33	53,97	46,03	25,00	75,00	28,57	71,43	100,00	0,00	30,77	69,23	75,00	25,00	57,14	42,86
Maret	36,80	63,20	27,91	72,09	45,83	54,17	66,67	33,33	77,78	22,22	34,38	65,62	80,00	20,00	57,14	42,86
April	35,38	64,62	64,84	35,16	8,51	91,49	100,00	0,00	100,00	0,00	46,43	53,57	100,00	0,00	9,10	90,90
Mei	58,18	41,82	59,32	40,68	45,76	54,24	23,08	76,92	100,00	0,00	74,55	25,45	55,56	44,44	19,05	80,95
Juni	35,29	64,71	61,82	38,18	31,82	68,18	90,00	10,00	71,43	28,57	76,19	23,81	76,92	23,08	0,00	100,00
Juli	45,12	54,88	70,83	29,17	39,39	60,61	100,00	0,00	18,75	81,25	71,43	28,57	33,33	66,67	53,57	46,43
Agustus	53,13	46,87	56,67	43,33	68,06	31,94	50,00	50,00	43,75	56,25	50,72	49,28	46,67	53,33	84,00	16,00
September	64,21	35,79	44,21	55,79	59,74	40,26	85,71	14,29	50,00	50,00	59,46	40,54	81,48	18,52	50,00	50,00
Oktober	77,55	22,45	46,61	53,39	60,20	39,80	100,00	0,00	50,00	50,00	57,89	42,11	79,59	20,41	71,79	28,21
November	76,92	23,08	62,96	37,04	47,50	52,50	66,67	33,33	14,29	85,71	75,86	24,14	41,18	58,82	75,00	25,00
Desember	50,00	50,00	65,63	34,37	78,95	21,05	16,67	83,33	0,00	100,00	34,04	65,96	50,00	50,00	73,33	26,67



Gambar 2. Persentase Sentimen Positif dan Negatif di twitter tiap Bulan pada Tahun 2018 di Kota Semarang

C. Tingkat Kematangan Keamanan Kota berdasar Persepsi Keamanan Masyarakat

Hasil penelitian persepsi masyarakat melalui analisis sentimen twitter terhadap *tweet* tentang keamanan kota di 25 kota Indonesia didapatkan bahwa *score* tingkat kematangan keamanan kota

berdasar persepsi masyarakat 25 kota di Indonesia di tunjukkan seperti pada tabel 11. Dari tabel 11 di tunjukkan bahwa terdapat 4 kota dengan tingkat kematangan keamanannya berada pada tingkat *integrative* (*score* antara 60 sampai dengan 79 pada *GSCM Maturity Level*) yaitu Tangerang, Kediri, Probolinggo dan Parepare, sedangkan 21 kota lainnya berada pada tingkat kematangan keamanan *Scattered* (*score* antara 40 sampai dengan 59). Dari tabel 11 ditunjukkan juga bahwa rata-rata tingkat kematangan keamanan untuk kategori kota besar adalah tingkat *Scattered* dengan *score* 55,41 , untuk kategori kota ukuran sedang berada pada tingkat *Scattered* dengan *score* 55,48 , dan untuk kategori kota ukuran kecil juga berada pada tingkat *Scattered* dengan *score* 53,70.

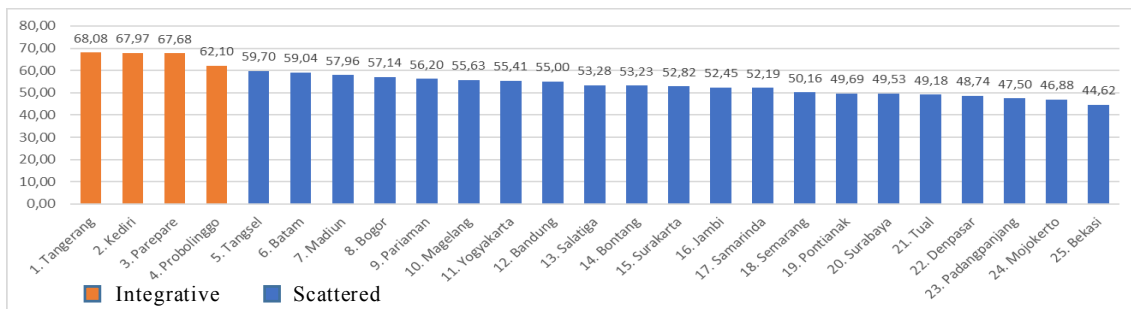
Tabel 11.

Score tingkat keamanan 25 kota Indonesia berdasar prediksi sentimen positif keamanan kota di twitter

Kategori Kota	Nama Kota	Sentimen Positif Keamanan Kota	Tingkat Kematangan Keamanan Kota	Rata-rata Tingkat Kematangan Kategori Kota
Kota Besar	Surabaya	49,53	<i>Scattered</i>	<i>Scattered</i> (55,41)
	Bekasi	44,62	<i>Scattered</i>	
	Bandung	55,00	<i>Scattered</i>	
	Semarang	50,16	<i>Scattered</i>	
	Tangsel	59,70	<i>Scattered</i>	
	Batam	59,04	<i>Scattered</i>	
	Bogor	57,14	<i>Scattered</i>	
	Tangerang	68,08	<i>Integrative</i>	
Kota Sengah	Denpasar	48,74	<i>Scattered</i>	<i>Scattered</i> (55,48)
	Jambi	52,45	<i>Scattered</i>	
	Yogyakarta	55,41	<i>Scattered</i>	
	Kediri	67,97	<i>Integrative</i>	
	Surakarta	52,82	<i>Scattered</i>	
	Madiun	57,96	<i>Scattered</i>	
	Pontianak	49,69	<i>Scattered</i>	
	Probolinggo	62,10	<i>Integrative</i>	
	Samarinda	52,19	<i>Scattered</i>	
Kota Kecil	Bontang	53,23	<i>Scattered</i>	<i>Scattered</i> (53,70)
	Magelang	55,63	<i>Scattered</i>	
	Mojokerto	46,88	<i>Scattered</i>	
	Tual	49,18	<i>Scattered</i>	
	Padangpanjang	47,50	<i>Scattered</i>	
	Pariaman	56,20	<i>Scattered</i>	
	Parepare	67,68	<i>Integrative</i>	
	Salatiga	53,28	<i>Scattered</i>	

Dari tabel 11 ini jika di gambarkan dalam bentuk rating tingkat keamanan kota berdasar persepsi masyarakat di 25 kota di Indonesia seperti di tunjukkan pada gambar 3 dengan kota

Tangerang sebagai kota terbaik yang memperoleh persentase sentimen positif keamanan kotanya sebesar 68,08 % , sementara urutan terbawah adalah kota Bekasi yang memperoleh persentase sentimen positif paling kecil dibanding kota lainnya yaitu sebesar 44,62 %.



Gambar 3. Rating tingkat keamanan kota (sentimen positif keamanan kota di twitter) di 25 kota Indonesia

V. EVALUASI

A. Evaluasi Model Klasifikasi

Validasi dari model klasifikasi yang digunakan penulis adalah menggunakan metode *10 fold cross validation* sedangkan kinerja algoritma model klasifikasinya diukur dengan *confusion matrix*. Pengukuran ini dilakukan dengan menghitung rata-rata akurasi, presisi dan *recall* dari 10 iterasi pada *10 fold cross validation*. Hasil pengukuran untuk akurasi prediksinya adalah 80,10% , presisi sebesar 81,10% serta *recall* sebesar 82,62%.

B. Evaluasi Hasil Penelitian

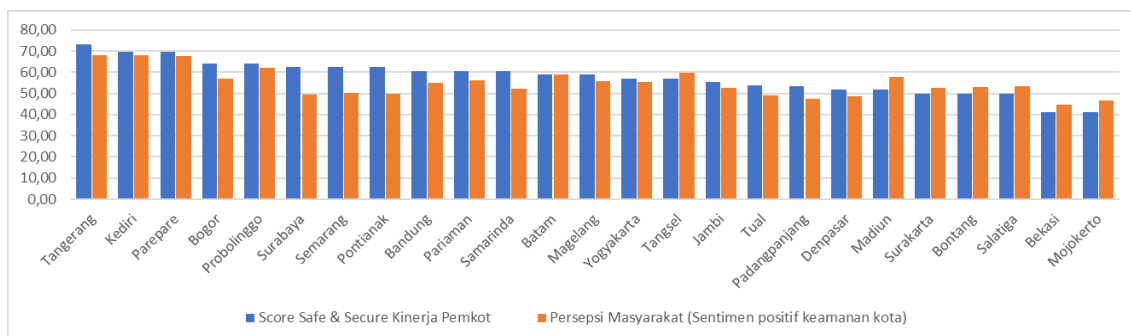
Evaluasi hasil penelitian dilakukan dengan membandingkan hasil penelitian yang dilakukan penulis dengan hasil survei RKCI tentang tingkat keamanan kota. Pada tabel 12 ditunjukkan perbandingan *score* tingkat keamanan 25 kota di Indonesia berdasar penelitian penulis tentang persepsi masyarakat (sentimen positif keamanan kota) dengan hasil survei RKCI tahun 2019 yang merupakan kinerja Pemkot di bidang keamanan dan kebencanaan selama tahun 2018. Berdasar data tabel 12 ini dapat digambarkan dalam bentuk *chart* seperti pada gambar 4 terlihat bahwa *score* sentimen positif tidak jauh berbeda dengan *score* hasil survei RKCI, sehingga dapat dikatakan bahwa secara umum persepsi masyarakat (sentimen positif keamanan kota) berbanding lurus dengan kinerja Pemkot. Kalau kinerja Pemkot baik maka persepsi masyarakat juga baik.

Tabel 12.

Perbandingan score tingkat keamanan 25 kota di Indonesia berdasar persepsi masyarakat (sentimen positif keamanan kota) dari hasil penelitian dengan hasil survei RKCI bidang keamanan dan kebencanaan.

No	Kategori Kota	Nama Kota	Score Hasil Survei RKCI (bidang keamanan & kebencanaan)	Score Hasil Penelitian (Sentimen Positif Twitter)
1	Kota Besar	Surabaya	62,50	49,53
2		Bekasi	41,07	44,62
3		Bandung	60,71	55,00
4		Semarang	62,50	50,16
5		Tangsel	57,14	59,70
6		Batam	58,93	59,04

7		Bogor	64,29	57,14
8		Tangerang	73,21	68,08
9	Kota Sedang	Denpasar	51,79	48,74
10		Jambi	55,36	52,45
11		Yogyakarta	57,14	55,41
12		Kediri	69,64	67,97
13		Surakarta	50,00	52,82
14		Madiun	51,79	57,96
15		Pontianak	62,50	49,69
16		Probolinggo	64,29	62,10
17		Samarinda	60,71	52,19
18		Kota Kecil	Bontang	50,00
19	Magelang		58,93	55,63
20	Mojokerto		41,07	46,88
21	Tual		53,85	49,18
22	Padangpanjang		53,57	47,50
23	Pariaman		60,71	56,20
24	Parepare		69,64	67,68
25	Salatiga		50,00	53,28



Gambar 4. Perbandingan *score* tingkat keamanan 25 kota di Indonesia berdasar persepsi masyarakat (sentimen positif keamanan kota) dari hasil penelitian dengan hasil survei RKCI tahun 2019 bidang keamanan dan kebencanaan

Namun ada 3 kota yang perbedaannya cukup mencolok yaitu Surabaya, Semarang dan Pontianak. Ketiga kota ini *score* persepsi masyarakatnya lebih rendah di atas 10% dibanding *score* hasil survei RKCI. Untuk kota Surabaya ini disebabkan ada kejadian terorisme yaitu serangan bom di 3 gereja pada pertengahan tahun 2018 yang menyebabkan percakapan *tweet* tentang “terorisme” menduduki peringkat tertinggi mencapai 35,74% selama tahun 2018, sehingga hal ini sangat mempengaruhi persentase *score* sentimen negatif *tweet*nya. Sedangkan kota Semarang dipengaruhi oleh banyaknya percakapan negatif tentang “banjir” di Semarang yang jumlah

tweetnya mencapai 48,28% selama tahun 2018. Demikian halnya dengan kota Pontianak, *score* lebih rendahnya dipengaruhi oleh banyaknya *tweet* tentang “kebakaran” hutan dan lahan yang mencapai 35,50% selama tahun 2018.

VI. KESIMPULAN

Berdasarkan hasil penelitian dan evaluasi yang telah dilakukan, terdapat beberapa kesimpulan yang dapat diperoleh yaitu sebagai berikut.

1. Hasil dari model prediksi analisis sentimen persepsi masyarakat dalam penelitian ini secara umum tidak jauh berbeda dengan hasil survei RKCI 2019.
2. Hasil pengukuran kinerja dari model prediksi analisis sentimen persepsi masyarakat dalam penelitian ini adalah untuk nilai akurasi sebesar 80,10% sedangkan nilai presisi sebesar 81,10% dan nilai recall sebesar 82,62%.
3. Berdasarkan hasil penelitian, rata-rata *score* tingkat keamanan untuk kategori kota besar di Indonesia dalam penelitian ini rata-ratanya 55,41 (*Scattered*), untuk rata-rata kota kategori sedang adalah 55,48 (*Scattered*) dan kategori kota kecil sebesar 53,70 (*Scattered*).
4. Berdasarkan hasil penelitian 25 kota Indonesia ini didapatkan bahwa 4 kota tingkat kematangan keamanan kotanya berada pada tingkat *integrative* yaitu Tangerang (*score* 68,08), Kediri (*score* 67,97), Parepare (*score* 67,68), Probolinggo (*score* 62,10). Sedangkan 21 kota lain tingkat kematangannya berada pada tingkat *scattered* dengan *score* antara 40 sampai 59.

DAFTAR PUSTAKA

- [1] H.S. Firmansyah, S.H. Supangkat, A.A Arman dan R. Adhitya : Searching Smart City in Indonesia Through Maturity Model Analysis, *International Conference on ICT for Smart Society (ICISS) 2017*.
- [2] M. Lacinak, dan J. Ristvej : Smart city, Safety and Security, *TRANSCOM 2017 International scientific conference on sustainable, modern and safe transport*.
- [3] Garuda Smart City Model (GSCM), Smart City and Community Innovation Center (SCCIC) Institut Teknologi Bandung.
- [4] J.M. Soler, F. Cuartero, dan M. Roblizo : Twitter as a tool for predicting elections results. *IEEE/ACM International Conference on advances in social networks analysis and mining (ASONAM)*, 2012. Piscataway: IEEE; pp. 1194–1200.
- [5] W. Budiharto, dan M. Meiliana : Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis, *Journal of Big data*. 2018.
- [6] R. Fieldman, dan J. Sanger : *The Text Mining Handbook*, Cambridge: Cambridge University Press, 2006.
- [7] B. Liu : Sentiment Analysis and Opinion Mining, *A Review Article on Synthesis Lectures on Human Language Technologies*, Vol. 5, No-1, pp. 1-167, April- 2012.
- [8] J. Zhu, E. Zhuang, J. Fu, dan J. Baranowski : A Framework-Based Approach to Utility Big data analytics, *IEEE Transaction on Power Systems*, Vol.31, No.3, 2016.
- [9] B. Pang, dan L. Lee : A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, *ACL 2014*.
- [10] A.B. Ghulam : Sentiments Analysis for Governor of East Java 2018 in Twitter, *Journal Publications & Informatics Engineering Research* Volume 3, Number 2, April 2019.
- [11] E. Prasetyo, *Data mining : Konsep Dan Aplikasi Menggunakan Matlab*, 1st ed. Yogyakarta : Andi, 2012.
- [12] M.W. Wijaya : *Teknologi Big data : Sistem Canggih dibalik Google, Yahoo, Facebook, IBM*, Edisi kedua (2019).
- [13] H.S Priyanka : Classification Model To Determine The Polarity Of Movie Review Using Logistic Regression, *International Research Journal of Computer Science (IRJCS)* , 2019.