Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode *Support Vector Machine*

Anita Novantirani[1], Mira Kania Sabariah S.T., M.T[2], Veronikha Effendy, ST.,M.T[3]
Program Studi Teknik Informatika, Fakultas Informatika, Universitas Telkom, Bandung
[1]anitnitnit@gmail.com, [2]mira_ljuana@yahoo.com, [3] veffendy@gmail.com

Abstrak

Kemacetan yang terjadi di kota besar di Indonesia diakibatkan oleh menjamurnya penggunaan kendaraan pribadi. Solusi untuk mengurangi kemacetan tersebut adalah dengan peningkatan penggunaan transportasi umum darat dalam kota, yang nyatanya masih belum begitu diminati masyarakat. Sebagian masyarakat menyampaikan pendapat dan opininya mengenai penggunaan transportasi umum dalam kota melalui *Twitter*. Opini tersebut dapat dimanfaatkan sebagai bahan analisis sentimen untuk mengetahui penilaian pelayanan transportasi umum darat dalam kota apakah positif atau negatif, serta mengetahui faktor opini apa yang sering muncul. Hasil dari analisis sentimen tersebut dapat membantu dalam penilaian dan evaluasi terhadap penggunaan transportasi umum darat dalam kota. Dengan dilakukannya peningkatan fasilitas dan pelayanan berdasarkan hasil analisis sentimen, maka diharapkan masyarakat akan beralih menggunakan transportasi umum darat dalam kota, yang tentunya akan mengurangi kemacetan. Analisis sentimen dengan metode *Support Vector Machine* (SVM) dilakukan dengan pengujian terhadap komposisi data yang bervariasi. Dari hasil pengujian untuk kasus pada penelitian ini didapatkan bahwa SVM dapat diimplementasikan dengan nilai akurasi mencapai 78,12%. Variabel yang berpengaruh terhadap akurasi adalah jumlah data, perbandingan jumlah data latih dan uji, serta perbandingan jumlah data positif dan negatif yang digunakan.

Kata Kunci : Transportasi Umum Darat Dalam Kota, Analisis Sentimen, *Twitter*, *Support Vector Machine*

1. Pendahuluan

Kemacetan sering terjadi di berbagai wilayah, khususnya wilayah kota besar di Indonesia. Ketua Dewan Transportasi Kota Jakarta, Azas Tigor Nainggolan, beranggapan bahwa kemacetan diakibatkan oleh menjamurnya penggunaan kendaraan-kendaraan pribadi [1]. Peningkatan penggunaan angkutan umum dapat menjadi solusi kunci permasalahan kemacetan [2], yang jika berhasil akan membuat pengguna kendaraan pribadi beralih ke transportasi umum dan akan mengurangi jumlah volume kendaraan di jalan yang tentunya akan mengurangi kemacetan. Namun kebanyakan masyarakat masih enggan menggunakan transportasi umum dan lebih memilih menggunakan kendaraan pribadi. Keengganan tersebut didasari oleh beberapa faktor, diantaranya mengenai waktu perjalanan, ongkos, keselamatan dan keamanan, serta kesenangan dan kenyamanan pengguna transportasi umum itu sendiri [3]. Hal ini perlu ditilik lebih lanjut dengan mengetahui pendapat masyarakat tentang apa yang dirasakan saat menggunakan transportasi umum darat dalam kota di Indonesia.

Masyarakat memberikan pendapat dan opininya melalui berbagai media. Salah satu media yang efektif untuk menampung opini penggunaan transportasi umum ini adalah *Twitter*, yang termasuk cepat dalam mengabarkan pengalaman yang dirasakan oleh masyarakat sebagai bahan evaluasi untuk pihak trasnsportasi umum. Selain itu *Twitter* sendiri merupakan salah satu media sosial yang akrab digunakan oleh masyarakat Indonesia, yang tentunya akan memudahkan untuk pengumpulan opini dibandingkan dengan melakukan *survey* ataupun penyebaran kuisioner. Berdasarkan penelitian Semiocast, lembaga riset media sosial yang berpusat di Paris, Prancis, jumlah pemilik akun *Twitter* di Indonesia merupakan yang terbesar kelima di dunia, dan berada pada posisi ketiga negara yang paling aktif mengirim pesan *Twitter* (*tweet*) perhari [4]. Dengan banyaknya pengguna disertai kemudahan pengaksesan *Twitter* dalam penyampaian opini, maka akan banyak data opini terkumpul yang kemudian bisa menjadi peluang untuk dimanfaatkan sebagai bahan penilaian dan evaluasi atas penggunaan transportasi umum darat dalam kota di Indonesia.

Untuk menghasilkan informasi dari data opini yang ada, dilakukan pengolahan data dengan analisis

sentimen yang dapat menjadi solusi permasalahan yang telah dipaparkan sebelumnya, yaitu dengan memisahkan opini ke dalam kelas bersentimen positif atau negatif, serta dilakukan pengambilan kesimpulan faktor apa yang sering dibahas dalam opini-opini tersebut. Sebelumnya telah dilakukan penelitian analisis sentimen terhadap *Twitter* dengan berbagai metode seperti *Naïve Bayes Classification*, *Maximum Entropy*, ataupun *Support Vector Machine* [5][6] . Didapatkan kesimpulan bahwa penggunaan metode *Support Vector Machine* (SVM) memberikan hasil paling baik dibandingkan metode lainnya, yaitu dengan tingkat keakuratan hingga 82,2% [6]. Metode SVM sendiri dapat digunakan untuk mengklasifikasikan data opini berdasarkan atribut penilaian yang dimiliki untuk dapat dipisahkan apakah opini termasuk ke dalam kelas bersentimen positif atau negatif. Oleh karena itu, dalam penelitian ini digunakan metode SVM yang diharapkan sesuai dengan permasalahan. Dengan dilakukannya analisis sentimen ini, diharapkan dapat diketahui permasalahan yang terdapat pada penggunaan transportasi umum darat dalam kota di Indonesia, sehingga dapat dijadikan saran dan masukan untuk peningkatan penggunaan transportasi umum yang diharapkan juga dapat membantu mengurangi kemacetan.

2. Landasan Teori

2.1 Twitter

Twitter adalah situs web dimiliki dan dioperasikan oleh Twitter, Inc., yang 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.. Twitter memiliki karakteristik dan format penulisan yang unik dengan simbol ataupun aturan khusus.Pesan dalam Twitter dikenal dengan sebutan *tweet*. [7]

2.2 Analisis Sentimen

Analisis sentimen, yang disebut juga dengan *opinion mining*, merupakan salah satu cabang ilmu dari data mining yang bertujuan untuk menganalisis, memahami, mengolah, dan mengekstrak data tekstual yang berupa opini terhadap entitas seperti produk, servis, organisasi, individu, dan topik tertentu [8]. Analisis ini digunakan untuk mendapatkan suatu informasi tertentu dari suatu kumpulan data yang ada. Analisis sentimen berfokus pada pengolahan opini yang mengandung polaritas, yaitu memiliki nilai sentimen positif ataupun negatif.

2.3 Transportasi Umum Darat Dalam Kota

Transportasi umum darat dalam kota, atau disebut juga angkutan kota, adalah angkutan dari suatu tempat ke tempat lain dalam wilayah kota dengan menggunakan mobil bus dan/atau mobil penumpang umum yang terkait dalam trayek tetap dan teratur[10] [10].

Ketersediaannya transportasi umum yang baik tentunnya akan memberikan manfaat bagi masyarakat secara umum. Masyrakat akan terbantu jika ingin berpergian ke suatu tempat tujuan. Apabila transportasi umum dijaga dan dirawat dengan baik, tentu akan memberikan dampak baik, yaitu masyrakat akan lebih memilih menggunakan transportasi umum dibandingkan kendaraat pribadi.

2.4 Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi dengan menggunakan *machine learning* (*supervised learning*) yang memprediksi kelas berdasarkan model atau pola dari hasil proses *training*. Klasifikasi dilakukan dengan mencari *hyperplane* atau garis pembatas (*decision boundary*) yang memisahkan antara suatu kelas dengan kelas lain, yang dalam kasus ini garis tersebut berperan memisahkan tweet bersentimen positif (berlabel +1) dengan tweet bersentimen negatif (berlabel -1). SVM melakukan pencarian nilai *hyperplane* dengan menggunakan *support vector* dan nilai *margin* [9]. Dalam penelitian ini, data inputan yang memiliki representasi vektor didapatkan dari proses pembobotan. Dengan dilakukannyan training pada klasifikasi SVM, maka akan menghasilkan sebuah nilai atau pola yang akan digunakan pada proses testing untuk proses testing SVM, yang bertujuan memberi label sentimen pada tweet [10].

2.5 Validasi dan Evaluasi

Validasi yang digunakan dalam penelitian ini adalah *k-fold cross validation*. Dalam *k-fold cross validation*, data awal dipartisi secara acak menjadi sejumlah k subset (fold), yaitu D_1, D_2, …, D_k, yang masing-masingnya berukuran sama. Proses *training* dan *testing* dilakukan sebanyak k kali. Dalam iterasi ke-i, data subset D_i digunakan sebagai dataset *testing*, sedangkan data lainnya digunakan dalam proses

training untuk mendapatkan suatu model klasifikasi yang nantinya akan digunakan dalam proses *testing*.

Evaluasi performansi dilakukan untuk menguji hasil dari klasifikasi dengan mengukur nilai performansi dari sistem yang telah dibuat. Parameter pengujian yang digunakan untuk evaluasi yaitu akurasi yang perhitungannya diperoleh dari tabel *coincidence matrix* (matrik klasifikasi atau tabel kontigensi)

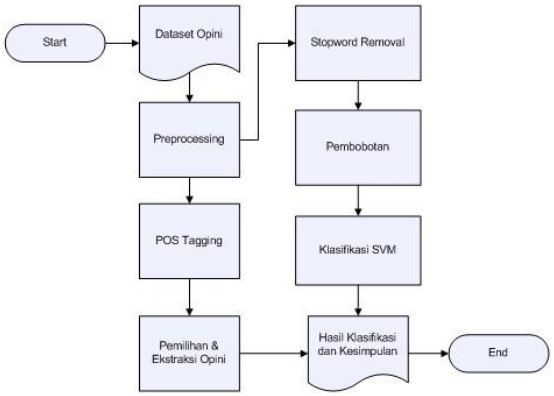
Tabel 2-1 Coincidence Matrix [11]

		True Class	
		Positive	**Negative**
Predicted Class	**Positive**	**True Positive Count (TP)**	**False Positive Count (FP)**
	Negative	**False Negative Count (FN)**	**True Negative Count (TN)**

3. Perancangan Sistem

Secara umum, sistem yang akan dibuat pada penelitian ini adalah sistem yang dapat menganalisis sentiment mengenai penggunaan transportasi umum darat dalam kota di Indonesia. Sistem akan memberikan hasil analisis sentiment opini pada suatu transportasi umum yang terklasifikasi sebagai sentiment positif atau negatif, serta akan disampaikan juga faktor opini yang paling sering dibicarakan pada tiap transportasi umum tersebut.

Gambaran umum sistem yang akan dibuat dalam penelitian tugas akhir ini adalah sebagai berikut:



Gambar 3-1 Gambaran Umum Sistem

Skenario keseluruhan sistem adalah sebagai berikut :

1. Sistem menerima inputan berupa *dataset twitter* bertipe file *excel*. Data tersebut di ambil dari berdasarkan input *query* menggunakan *tools* NodeXL. Dataset ini dikelompokkan menjadi empat sesuai jenis transportasi umum darat dalam kota.
2. Dilakukan *preprocessing* terhadap dataset dengan tujuan untuk membuat bentuk dataset sesuai dengan kebutuhan klasifikasi dan memudahkan pemrosesan pada sistem. *Preprocessing* terdiri dari subproses data *cleansing*, yang berupa *case folding* dan *remove noise*, subproses tokenisasi, dan subproses normalisasi kata
3. Pencarian faktor opini dilakukan dua proses, yaitu pemberian tag kata dengan menggunaka *POS tagging*. Kemudian berdasarkan tag kata tersebut, kata akan dipilih untuk dilakukan pemilihan ekstraksi fitur opini
4. Untuk pencarian hasil klasifikasi sentimen dilakukan proses *stopword removal*, yaitu untuk menghapus kata yang sering muncul tetapi tidak memiliki makna spesifik. Kemudian dilakukan pembobotan kata berdasarkan jumlah kemunculan dalam dokumen, sehingga dokumen dapat direpresentasikan dalam vektor. Fitur pembobotan yang digunakan yaitu unigram, dan metode pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF). Proses selanjutnya adalah pengklasifikasian sentimen data dengan menggunakan metode *Support Vector Machine* (SVM).

.

4. Pengujian dan Analisis

4.1 Dataset

Dataset yang digunakan dalam penelitian ini merupakan opini berbahasa Indonesia mengenai transportasi umum darat dalam kota yang berasal dari Twitter. Dataset berupa opini twitter yang pada dasarnya merupakan ungkapan masyarakat mengenai suatu hal melalui media sosial. Pada media sosial twitter sendiri lebih sering digunakan bahasa yang tidak formal, menggunakan tata bahasa yang cenderung bebas, dan tidak sedikit kata tidak baku yang muncul dalam *tweet*.

Dataset yang digunakan berfokus pada opini berbahasa Indonesia yang membahas tentang penggunaan transportasi umum darat khusus dalam

kota dengan memilih 4 sampel kendaraan dengan rincian jumlah dataset sebagai berikut:

Tabel 4-1 Rincian Dataset

Dataset	Rincian Data
Angkot	272 Kalimat Opini
Kopaja	184 Kalimat Opini
Metro Mini	264 Kalimat Opini
Transjakarta	418 Kalimat Opini

4.2 Skenario Pengujian

Untuk mencapai tujuan penelitian tugas akhir ini, maka dirancang tiga skenario pengujian. Skenario pengujian tersebut adalah sebagai berikut:

1. **Pengujian dataset dengan parameter komposisi data testing dan training berbeda**

Pengujian dilakukan dengan pembagian data training dan data testing dengan komposisi berdasarkan pembagian data pada k-fold cross validation dengan nilai k=2 sampai dengan k=10. Pengujian ini bertujuan untuk mengetahui nilai k-fold paling optimal dalam proses klasifikasi opini pada kasus penelitian ini.

2. **Pengujian dataset dengan parameter komposisi data positif dan data negatif berbeda**

Pengujian dilakukan dengan penggunaan jumlah dokumen yang tidak seimbang antara data positif dan data negatif, yaitu dengan jumlah dokumen positif yang lebih banyak dari dokumen negative, begitu pula sabaliknya. Pengujian dilakukan beberapa kali dengan merandom dataset dengan maksud memperhatikan tingkat fairness dari hasil pengujian data.

Pada penelitian sebelumnya [12] dikatakan bahwa komposisi data positif dan negatif perlu seimbang untuk menghindari adanya bias atau pembelokan bobot kata dengan jumlah kemunculan yang tinggi pada pada data negatif saja atau positif saja. Oleh karena itu pengujian ini dilakukan untuk membuktikan apakah keadaan tersebut juga berlaku pada data penelitian ini.

Pengujian ini akan dilakukan pada data yang merupakan data yang memiliki akurasi terbaik dari hasil pengujian skenario 1.

3. **Pengujian faktor opini**

Pengujian ini dilakukan untuk mengetahui faktor opini yang paling mendominasi dari tiap jenis dataset transportasi. Pengujian dilakukan dengan memanfaatkan label kelas kata hasil POS tagging, kemudian dilakukan ekstraksi opini dengan menseleksi kata dengan kelas kata *adjective*, *adverb*, dan *verb*. Dari sekian banyak faktor opini tersebut, akan dimabil sepuluh terbesarnya.

4.3 Analisa Hasil Pengujian

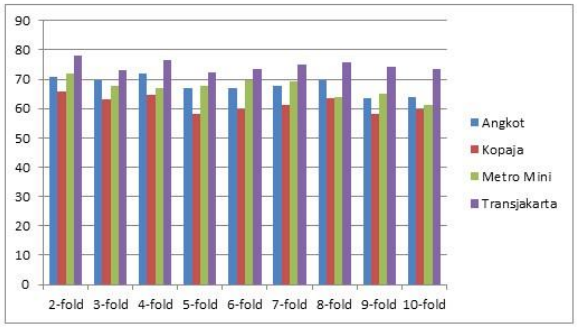
Pada bagian ini diperlihatkan hasil pengujian yang telah dilakukan beserta analisis terhadap hasil yang telah didapatkan dari proses pengujian.

4.3.1 Analisis Akurasi terhadap Komposisi Data Uji dan Data Latih Berbeda

Setelah dilakukan pengujian terhadap penggunaan komposisi data latih dan data uji berbeda berdasarkan pembagian data dengan validasi *k-fold cross validation*, maka dihasilkan akurasi seperti pada tabel 4-2. Dari pengujian tersebut dapat diketahui komposisi data latih dan data uji yang memberikan hasil akurasi paling optimal. Pada dataset Angkot, komposisi data paling optimal yaitu data latih berbanding data uji 3:1 (4-fold), sedangkan untuk dataset Kopaja, Metro Mini, dan Transjakarta didapatkan bahwa komposisi paling optimal 1:1 (2-fold).

Tabel 4-2 Rata-rata Akurasi Skenario Pengujian 1

Fold	Jenis Transportasi Umum			
	Angkot	Kopaja	Metro Mini	Transjakarta
2	70.9558824	**65.7608696**	**71.969697**	**78.125**
3	69.5249597	63.046595	67.8030303	73.26777
4	**72.0588235**	64.673913	67.04545455	76.68269
5	66.9708995	58.2748538	67.77777778	72.14286
6	66.8807642	59.7569444	69.6969697	73.51541
7	67.612782	61.4227877	69.4235589	74.97457
8	69.4852941	63.5416667	63.9476103	75.72115
9	63.7037037	58.1818182	65.1322751	74.28543
10	64.09340659	59.7777778	61.07142857	73.60714



Gambar 4-1 Grafik akurasi cross validation

Tinggi rendahnya nilai akurasi untuk pengujian klasifikasi tersebut dipengaruhi dari perbedaan porsi data latih dan data uji tersebut. Seperti yang terlihat pada tabel 4-2 dan gambar 4-1, semakin besar nilai

fold atau semakin banyak jumlah data latih, nilai akurasi yang dihasilkan cenderung menurun.

Berdasarkan hasil pengujian, dataset Transjakarta menghasilkan akurasi yang lebih baik dibandingkan dataset lainnya, dengan akurasi mencapai 78,125%. Hal ini dikarenakan dataset Transjakarta memiliki jumlah yang lebih banyak dibandingkan dataset lain sehingga keberagaman data semakin tinggi yang berpengaruh pada model yang dihasilkan dari proses training. Begitu juga sebaliknya yang terjadi pada dataset Kopaja yang memiliki jumlah terkecil, menghasilkan akurasi di bawah tiga dataset lainnya.

Dari keseluruhan hasil pengujian, rata-rata akurasi berada di kisaran 67,83%. Hasil akurasi tersebut disebabkan oleh penggunaan dataset berbahasa Indonesia, yang memiliki tata bahasa yang lebih kompleks dibandingkan dengan tata bahasa Inggris yang memiliki aturan khusus untuk penulisan pola kali tertentu. Hal ini juga didukung oleh penelitian sebelumnya [13], yang melakukan pengujian klasifikasi SVM terhadap data berbahasa Indonesia dan data berbahasa Inggris, yang menghasilkan nilai akurasi data berbahasa Indonesia yang lebih rendah dibandingkan data berbahasa Inggris.

Selain dikarenakan lebih kompleksnya tata bahasa Indonesia, bentuk dataset yang bersal dari Twitter ini juga mempengaruhi hasil akurasi karena data Twitter lebih cenderung menggunakan tata bahasa yang tidak teratur dan terdapat banyak kata tidak baku. Proses pembersihan data dan pembakuan data dengan memanfaatkan kamus untuk membakukan kata telah dilakukan pada tahap *preprocessing*, namun penggunaan kamus yang berasal dari penelitian sebelumnya [14] tidak dapat memenuhi kebutuhan kamus dalam kasus ini karena kondisi data yang memang bertata bahasa tidak baku dan cenderung bebas atau tidak beraturan.

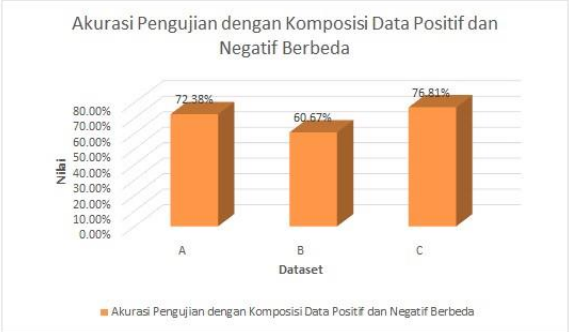
pengujian.

4.3.2 Analisis Akurasi terhadap Komposisi Data Positif dan Data Negatif

Data yang akan digunakan pada skenario pengujian ini adalah data yang memiliki hasil akurasi terbaik pada skenario pengujian 1, yaitu dataset Transjakarta. Pada pengujian ini, dataset tersebut dibagi menjadi tiga jenis dengan detail seperti pada tabel 4-3 berikut ini:

Tabel 4-3 Pembagian dataset skenario pengujian 2

No	Nama Dataset	Jumlah Tweet	Tweet Positif	Tweet Negatif
1	Dataset A	350	200	150
2	Dataset B	350	150	200
3	Dataset C	350	175	175



Gambar 4-2 Hasil rata-rata akurasi skenario pengujian 2

Dari gambar di atas dapat dilihat bahwa nilai akurasi untuk dataset C memiliki nilai akurasi paling baik diantara ketiga dataset yang ada, yaitu sebesar 76,81%. Sesuai dengan apa yang diungkapkan pada penelitian sebelumnya [12] bahwa dataset C yang memiliki komposisi data positif dan negatif yang seimbang akan memberikan hasil klasifikasi yang lebih baik karena menghindari terjadinya bias atau pembelokan bobot kata dengan jumlah kemunculan yang tinggi pada pada data negatif saja atau positif saja.

4.3.3 Analisis Faktor Opini

Setelah dilakukan pengujian terhadap faktor opini dengan menggunakan ekstraksi opini, maka diperoleh hasil faktor opini yang dominan berdasarkan tiap jenis transportasi seperti berikut:

Tabel 4-4 Faktor Opini Angkot

No.	Kata	Frekuensi	Jenis Sentimen
1	Mengetem	32	Negatif
2	Keren	25	Positif
3	Sialan	19	Negatif
4	Menunggu	18	Negatif
5	Lama	17	Negatif
6	Macet	17	Negatif
7	Tersedia	16	Positif
8	Hemat	16	Positif
9	Enak	15	Positif
10	Rokok	15	Negatif

Tabel 4-5 Faktor Opini Kopaja

No.	Kata	Frekuensi	Jenis Sentimen
1	Tersedia	19	Positif
2	Hemat	12	Positif
3	Macet	12	Negatif
4	Menunggu	11	Negatif
5	Keren	10	Positif
6	Duduk	9	Positif
7	Copet	8	Negatif
8	Mogok	6	Negatif
9	Penuh	6	Negatif
10	Rokok	6	Negatif

Tabel 4-6 Faktor Opini Metro Mini

No.	Kata	Frekuensi	Jenis Sentimen
1	Sembrono	30	Negatif
2	Baru	28	Positif
3	Tersedia	27	Positif
4	Macet	21	Negatif
5	Baik	16	Positif
6	Bagus	14	Positif
7	Keren	13	Positif
8	Nyaman	13	Positif
9	Senang	11	Positif
10	Copet	10	Negatif

Tabel 4-7 Faktor Opini Transjakarta

No.	Kata	Frekuensi	Jenis Sentimen
1	Baru	43	Positif
2	Lama (kedatangan)	43	Negatif
3	Dingin	41	Positif
4	Menunggu	41	Negatif
5	Cepat	38	Positif
6	Mogok	37	Negatif
7	Tersedia	35	Positif
8	Nyaman	23	Positif
9	Sepi	23	Positif
10	Duduk	20	Positif

Faktor yang mendasari keengganan masyarakat untuk beralih ke transportasi umum diantaranya mengenai waktu perjalanan, ongkos, keselamatan dan keamanan, serta kesenangan dan kenyamanan pengguna transportasi umum itu sendiri [3]. Dari hasil pengujian dan analisis yang telah dilakukan, didapatkan faktor opini yang memang sesuai dengan pernyataan tersebut, yaitu mengenai waktu perjalanan yang cenderung lebih banyak dikeluhkan lama dan macet pada Angkot, Kopaja, dan Metro Mini, namun ada juga yang berpendapat waktu perjalanan cepat khususnya pada Transjakarta. Hal yang juga berhubungan dengan waktu yaitu ketepatan waktu kedatangan transportasi, masih menjadi masalah karena masih banyak masyarakat yang mengeluhkan menunggu serta mengetem. Sedangkan untuk faktor ongkos tidak begitu dominan muncul pada hasil, ini menunjukkan bahwa masyarakat bisa menerima range ongkos yang ada saat ini karena hal tersebut tidak banyak dikeluhkan. Untuk faktor keselamatan dan keamanan terdapat opini tentang sikap sembrono atau ugal-ugalan, khususnya pada transportasi Metro Mini. Selain itu faktor keamanan yang juga menjadi perhatian adalah copet atau pencurian, hal ini dikeluhkan kususnya pada transportasi Metro Mini dan Kopaja. Faktor kesenangan dan kenyamanan pengguna transportasi umum juga menjadi perhatian. Pada Transjakarta terlihat bahwa banyak pengguna yang senang karena beberapa hal terkait kesenangan dan kenyamanan, yaitu terdapat faktor opini baru, nyaman, dingin, sepi, dan mendapatkan tempat duduk. Untuk Metro Mini terdapat faktor opini baik, bagus, keren, nyaman. Faktor kenyamanan yang diterima Kopaja relatif seimbang antara mendapatkan tempat duduk dan penuh armada.

Berdasarkan hasil analisis yang telah diuraikan di atas, dapat disimpulkan bahwa transportasi darat dalam kota memiliki penilaian yang beragam di mata masyarakat. Ada opini positif, da nada pula opini negatif bagi tiap jenis transportasi yang diteliti. Opini-opini positif tersebut dapat dimanfaatkan untuk dapat dilakukan ditingkatkan atau pun bisa diterapkan pada transportasi jenis lain, begitu pula dengan opini-opini negatif bisa dijadikan saran dan masukan bagi pihak terkait untuk dapat meningkatkan pelayanan transportasi umum, khususnya transportasi umum darat di dalam kota. Dengan peningkatan pelayanan tersebut diharapkan dapat membuat masyarakat mau beralih dari menggunakan kendaraan pribadi menjadi menggunakan transportasi umum, yang apabila terjadi juga diharapkan akan dapat membantu mengurangi kemacetan.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil pengujian dan analisis yang telah dilakukan pada bahasan sebelumnya, maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Analisis sentimen terhadap data Twitter mengenai penggunaan transportasi umum darat dalam kota dapat dilakukan dengan metode *Support Vector Machine*, dengan akurasi mencapai 78,12% pada dataset Transjakarta.
2. Hasil akurasi terhadap penggunaan metode *Support Vector Machine* dipengaruhi oleh beberapa hal, yaitu
   - Komposisi jumlah data training dan testing
   - Jumlah dataset yang digunakan
   - Komposisi jumlah data positif dan negatif
3. Faktor opini yang mendominasi dari hasil analisis sentiment terhadap penggunaan

transportasi darat dalam dalam kota diantaranya yaitu

- Angkot : mengetem, keren, sialan
- Kopaja : tersedia, hemat, macet
- Metro Mini : sembrono, baru, tersedia
- Transjakarta : baru, lama (kedatangan), dingin

5.2 Saran

Saran yang ingin disampaikan untuk pengembangan lebih lanjut Tugas Akhir ini adalah sebagai berikut :

1. Pengembangan sistem online yang dapat mengambil dataset dan menghasilkan sentiment secara realtime.
2. Perlu dibuat kamus khusus untuk penanganan kasus yang menggunakan data Twitter berbahasa Indonesia, yang cenderung menggunakan bahasa yang tidak baku, agar performansi menjadi lebih baik.
3. Data yang dihasilkan dari sistem yang berupa opini dan faktor penilaiannya dapat dimanfaatkan bagi pihak terkait untuk meningkatkan kinerja dalam pelayanan transportasi umum darat, khususnya dalam kota di Indonesia.
4. Untuk penanganan kasus klasifikasi dapat digunakan metode lain sesuai dengan kebutuhan data ataupun sesuai dengan kasus.

DAFTAR PUSTAKA

[1] A. Irmansyah. (2013, Aug.) Benahi Angkutan Umum Segera agar Tahun Depan Jakarta Tak Macet Total. [Online]. http://www.portalkbr.com/berita/perbincangan/2897988_4215.html

[2] Furqan. (2013, Apr.) Prof. Ade Sjafruddin: Angkutan Umum, Solusi Kunci Kemacetan Jakarta. [Online]. http://www.itb.ac.id/news/3899.xhtml

[3] L. Manheim, *Fundamental Transportation Systems Analysis*. The MIT Press, 1979, vol. I, Basic Concept.

[4] R. Asih. (2012, Feb.) Indonesia Pengguna Twitter Terbesar Kelima Dunia. [Online]. http://www.tempo.co/read/news/2012/02/02/072381323/

[5] M. Y. Nur and D. D. Santika, "Analisis Sentimen pada Dokumen Berbahasa Indonesia dengan Pendekatan Support Vector Machine," in *Konferensi Nasional Sistem dan Informatika*, Bali, 2011.

[6] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," Stanford, Project Report CS224N, 2009.

[7] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis," *HPL-2011-89*, vol. 89, Jun. 2011.

[8] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publisher, 2012.

[9] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques, Second Edition*, 2nd ed. Morgan Kaufmann, 2006.

[10] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing 10*, 2002, pp. 79-86.

[11] D. L. Olson and D. Delen, *Advanced Data Mining Techniques*, 1st ed. Heidelberg, Berlin: Springer, 2008.

[12] Zhang, Weishi, G. Ding, L. Chen, and C. Li, "Chinese Online Video Recomendation by Using Virtual Rating Predicted by Review Sentiment Classification," *IEEE International Conference on Data Mining Workshop*, 2010.

[13] N. W. S. Saraswati, "Text Mining dengan Metode Naïve Bayes Classifier dan Support Vector Machine untuk Sentimen Analysis," Program pasca Sarjana Universitas Udayana Thesis, 2011.

[14] A. F. Wicaksono and A. Purwarianti, "HMM Based Part-of-Speech Tagger for Bahasa Indonesia," *Proceedings of 4th International MALINDO (Malay – Indonesian Language)*, 2010.