

# Sentiment Analysis Masyarakat Pekalongan Terhadap Pembangunan Jalan Tol Pemalang-Batang Di Media Sosial

Ari Putra Wibowo, Eny Jumiati

Program Studi Teknik Informatika STMik Widya Pratama

Jl. Patriot 25 Pekalongan Telp (0285) 427816

Email: [ariputra.stmikwp@gmail.com](mailto:ariputra.stmikwp@gmail.com), [enyjumiati@stmik-wp.ac.id](mailto:enyjumiati@stmik-wp.ac.id)

## Abstrak

*Jalan tol Pemalang-Batang merupakan bagian jaringan jalan tol Trans Jawa yang dicanangkan oleh pemerintah dalam rangka mempercepat pembangunan infrastruktur nasional. Proses pembangunan jalan tol ini diprediksi memerlukan waktu penyelesaian selama 2 tahun. Sejak ditetapkannya awal pembangunan jalan tol Pemalang-Batang pada tahun 2016 oleh Presiden Indonesia Joko Widodo banyak opini masyarakat baik dunia nyata maupun di media sosial. Pada penelitian ini dilakukan sentiment analysis terkait pembangunana jalan tol Pemalang-Batang, data yang digunakan merupakan komentar dari postingan di group facebook. Analisis dilakukan dengan menggunakan metode naïve bayes, untuk pengujian performa naïve bayes menggunakan k fold cross validation. Dari hasil pengujian diperoleh nilai akurasi tertinggi sebesar 73,54% dengan nilai fold adalah 3.*

**Kata Kunci:** analisis sentimen, klasifikasi, naïve bayes

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Pembangunan jalan tol Pemalang-Batang merupakan proyek pembangunan yang sudah direncanakan sejak 20 tahun yang lalu. Jalan tol Pemalang-Batang merupakan bagian jaringan jalan tol Trans Jawa yang dicanangkan oleh pemerintah dalam rangka mempercepat pembangunan infrastruktur nasional (Jati 2016). Proses pembangunan jalan tol ini diprediksi memerlukan waktu penyelesaian selama 2 tahun. Sejak ditetapkannya awal pembangunan jalan tol Pemalang-Batang pada tahun 2016 oleh Presiden Indonesia Joko Widodo banyak opini masyarakat baik dunia nyata maupun di media sosial. Semua orang bebas berpendapat atau menyampaikan opininya terkait dengan pembangunan jalan tol Pemalang-Batang sehingga memunculkan banyak opini, tidak hanya opini yang bersifat positif namun banyak juga yang menyampaikan opini negatif.

Jejaring sosial atau media sosial adalah platform populer untuk interaksi, komunikasi, dan kolaborasi antar teman melalui koneksi internet (Wilson et al. 2012). Media sosial sekarang ini tidak hanya digunakan sebagai sarana pertemanan, mencari teman, tapi sudah banyak digunakan untuk kegiatan lain. Seperti sebagai sarana promosi produk, jual beli dan juga sebagai sarana

berbagi informasi (Buntoro 2016). Ulasan postingan atau tanggapan disampaikan melalui komentar di sosial media. Komentar tersebut mendeskripsikan tanggapan yang berbeda-beda dari setiap pengguna media sosial. Komentar yang berupa teks dapat dikumpulkan dan diolah dengan analisis sentimen. Analisis sentimen mempelajari opini yang mengungkapkan atau mengekspresikan pandangan positif atau negatif (Gusriani, Wardhani, and Zul 2016).

Analisis sentimen disebut juga opinion mining, adalah bidang ilmu yang menganalisa pendapat, sentimen, evaluasi, penilaian, sikap dan emosi publik terhadap entitas seperti produk, jasa, organisasi, individu, masalah, peristiwa, topik, dan atribut mereka (Zhang and Liu 2016). Beberapa penelitian terdahulu yang terkait dengan analisis sentimen pada media sosial antara lain adalah penelitian (Novantirani, Sabariah, and Effendy 2015) menganalisis sentimen pada teks twitter, untuk menganalisa penggunaan transportasi umum darat. Dari analisis sentimen tersebut dapat membantu dalam penilaian dan evaluasi terhadap penggunaan transportasi umum darat dalam kota. Penelitian (Buntoro 2016) menganalisis sentimen calon Gubernur DKI Jakarta 2017 di Twitter, Pada penelitian ini Analisis sentimen dilakukan untuk melihat sebuah opini seseorang yang ditujukan kepada calon

Gubernur DKI Jakarta 2017, opini itu bisa dimasukkan kategori opini positif, netral atau negatif. Besarnya sentimen yang ditujukan kepada calon Gubernur DKI Jakarta 2017 bisa dijadikan sebuah parameter kemenangan atau kekalahan seorang calon. Selanjutnya penelitian (Gusriani, Wardhani, and Zul 2016) melakukan analisis sentimen pada fanpage facebook toko online Berrybenka pada penelitian analisis sentimen ini dibahas bagaimana mengimplementasikan analisis sentimen untuk menentukan kecenderungan padangan publik terhadap toko online menggunakan metode klasifikasi naïve bayes.

Dalam penelitian ini Analisis sentimen dilakukan untuk melihat sebuah opini masyarakat terutama warga Pekalongan yang ditujukan pada pembangunan jalan tol Pemalang-Batang. Opini yang dibutuhkan untuk melakukan analisis berasal dari komentar postingan di group facebook Pekalongan Info. Pada penelitian ini akan dibahas bagaimana implementasi metode klasifikasi naïve bayes untuk menganalisis sentimen untuk menantukan kecenderungan opini masyarakat Pekalongan terhadap pembangunan jalan tol Pemalang-Batang. Naïve bayes merupakan metode pembelajaran mesin yang memiliki model dalam membentuk probabilitas dan peluang. Maka dari itu, naïve bayes akan menghitung probabilitas kemunculan kata yang mempresentasikan komentar berdasarkan kelas positif maupun negatif. Berdasarkan hasil klasifikasi dilakukan pengujian untuk mendapatkan nilai akurasi dari hasil prediksi analisis sentimen dengan metode naïve bayes (Gusriani, Wardhani, and Zul 2016).

## 1.2 Landasan Teori

### 1.2.1 Media Sosial

Jejaring sosial adalah platform populer untuk interaksi, komunikasi, dan kolaborasi antar teman melalui koneksi internet(Wilson et al. 2012).

Beberapa media sosial yang populer antara lain:

- a. Facebook
- b. Twitter
- c. LinkedIn
- d. Snapchat
- e. Instagram
- f. Pinterest
- g. Youtube

Dari beberapa sosial media yang sering digunakan Instagram adalah situs jejaring sosial yang paling banyak digunakan diikuti oleh Snapchat dan Facebook. Situs jejaring sosial yang paling jarang digunakan adalah LinkedIn dan Pinterest. Situs media sosial semakin disesuaikan untuk memenuhi kebutuhan target pasar tertentu. Memahami pola evolusi ini adalah kunci yang membuka platform media sosial mana yang akan terus digunakan oleh sebagian besar pengguna media (Knight-mccord et al. 2016).

### 1.2.2 Text Mining

Text mining merupakan bidang penelitian yang menarik informasi berharga dari teks yang tidak terstruktur (Dang and Ahmad 2014). Informasi biasanya diperoleh melalui peramalan pola dan kecenderungan pembelajaran pola statistik. Proses text mining yang khas meliputi kategorisasi teks, text clustering, ekstraksi konsep/entitas, produksi taksonomi granular, sentiment analysis, penyimpulan dokumen, dan pemodelan relasi entitas yaitu, pembelajaran hubungan antara entitas (Lidya, Sitompul, and Efendi 2015).

### 1.2.3 Sentiment analysis

Sentimen,analysis juga disebut opinion mining, adalah bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi orang-orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya (Zhang and Liu 2016).

### 1.2.4 Naïve Bayes

Naive bayes adalah algoritma yang paling efektif dan efisien dalam klasifikasi untuk data mining(Zaidi and Cerquides 2013), yang dibangun diatas jaringan Bayesian(Han, Kamber, and Pei 2012). Naive bayes merupakan salah satu algoritma machine learning yang menggunakan metode probabilitas. Probabilitas keanggotaan dapat diprediksi oleh pengklasifikasi dan proses didasarkan pada teorema bayes (Hall et al. 2011). Hasil klasifikasi dari bayes juga menunjukkan nilai akurasi dan akurasi yang tinggi jika diterapkan pada database besar (Han, Kamber, and Pei 2012).

Persamaan teorema bayes adalah:

$$P(X|H) = \frac{P(H|X)P(X)}{P(H)}$$

Dimana:

H : data dengan class yang belum diketahui  
X : hipotesis terhadap kelas H, merupakan suatu class

$P(X|H)$  : kemungkinan hipotesis X berdasar kondisi H

$P(H|X)$  : kemungkinan H berdasar kondisi pada hipotesis X

$P(X)$  : kemungkinan kelas X

$P(H)$  : kemungkinan dari H

Tahapan naive bayes :

- Menghitung peluang class/label
- Menghitung peluang masing-masing nilai data dengan masing-masing class
- Kalikan semua hasil dari masing-masing nilai data berdasarkan class yang sama
- Menghitung  $P(H|X)P(X)$
- Kesimpulan (bandingkan hasil dari masing-masing class).

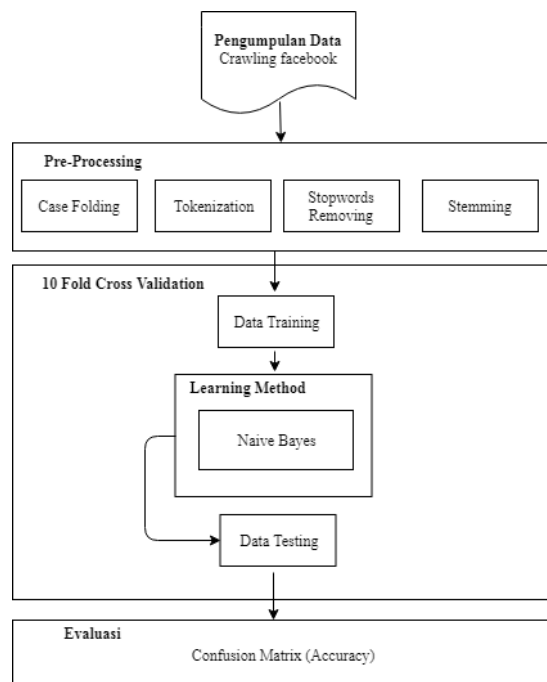
### 1.2.5 Cross Validation

Cross validation adalah metode statistik untuk mengevaluasi dan membandingkan algoritma pembelajaran (learning algorithms) dengan membagi data menjadi dua segmen, satu segmen digunakan untuk belajar atau data latih, dan yang lain digunakan untuk memvalidasi model (Mudry and Tjellström 2011). Dalam cross validation kumpulan pelatihan dan validasi harus cross-over berturut-turut sehingga setiap data memiliki kesempatan tervalidasi.

K-fold cross validation adalah teknik umum untuk memperkirakan kinerja pengklasifikasi. K-fold cross validation dilakukan dengan menggunakan kembali dataset yang sama, sehingga menghasilkan k perpecahan dari kumpulan data menjadi non-overlapping dengan proporsi pelatihan  $(k-1)/k$  dan  $1/k$  untuk pengujian (Singh 2010).

## 2. METODOLOGI PENELITIAN

Metode penelitian yang digunakan pada penelitian ini adalah metode eksperimen dan evaluasi. Tahapan penelitian ini adalah pengumpulan data, pre-processing, ekstraksi fitur, learning method dan validasi.



Gambar 1. Model yang Diusulkan

### 2.1 Metode Pengumpulan Data

Data yang digunakan pada penelitian ini diambil dari komentar facebook group pekalongan info terkait postingan pembangunan jalan tol Pemalang-Batang <https://www.facebook.com/groups/pekalonganinfo/permalink/1708150345946904/> yang diposting pada tanggal 7 April 2018. Postingan ini telah mendapat respon dari anggota group tidak kurang dari 5000 tanggapan, dan telah dibagikan sebanyak 70 kali serta lebih dari 700 komentar.

### 2.2 Pre-Processing

Tahapan pre-processing yang digunakan pada penelitian ini adalah case folding, tokenizing, stopwords serta stemming.

### 2.3 Validasi K-Fold Cross Validation

Validasi yang digunakan pada penelitian ini adalah k-fold cross validation. K-fold cross validation membagi data akan dibagi menjadi k bagian. Proses training dan testing dilakukan sebanyak k kali.

### 2.4 Evaluasi

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, presisi dan recall yang perhitungannya diperoleh dari tabel confusion matrix (matrik klasifikasi atau tabel kontigensi) (Novantirani, Sabariah, and Effendy 2015).

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

Gambar 2. Confusion matrix

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Proses eksperimen ini menggunakan aplikasi RapidMiner 7.6. Untuk pengujian model dilakukan menggunakan dataset yang diperoleh dari hasil crawling group facebook pekalongan info terkait pembangunan jalan tol Pemalang-Batang,

<https://www.facebook.com/groups/pekalonganinfo/permalink/1708150345946904>. Dari dataset yang sudah diperoleh, hanya digunakan 200 komentar terkait postingan tersebut yang terdiri dari 100 komentar positif dan 100 komentar negatif. Spesifikasi komputer yang digunakan untuk eksperimen ini dapat dilihat pada Tabel 1

Tabel 1 Spesifikasi komputer yang digunakan

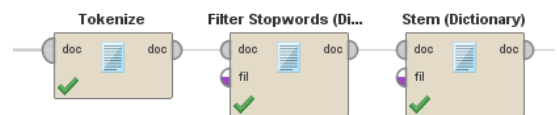
Processor	Intel(R) Core (TM) i5 CPU
Memory	6.00 GB
Harddisk	500 GB
Sistem Operasi	Windows 7 32 bit
Aplikasi	RapidMiner 7.6

1	Setuju... Buktinya pantura indramayu semenjak ada tol cipali sekarang jalan nya jadi bagus...nyaman dipakai gak kyak dulu
2	transportasi lancar perekonomian lancar gan harga menurun karna jarak yg ditempuh jd makin cepat. Jalan didesa jg aman dr mobil2 besar
3	Perekonomian indonesia akan maju, dan indonesia akan menjadi superpower di di asia, tapi sayang pribumi hanya menjadi pekerja, bukan pemilik
4	jika pribumi mau berusaha dia bisa jadi pemilik.. Sperti halnya ttg pendidikan, jika pribumi tdk mau belajar yg giat biar pintar dimasa yg akan datang dia punya ilmu untuk dikembangkan dan mnjadi motivasi untuk membangun sesuatu yang unik sehingga dia bisa jadi pemilik.. Semua bisa asal mau berusaha
5	setiap pembangunan pasti ada pengorbanan,nanti pada waktu proyek tol sudah selesai dan sudah diresmikan pasti jalan2 seluruh pekalongan yg terkena dampak proyek tol pasti akan di bagusin seperti semula bahkan lebih bagus lagi.
6	Apa yg kita kerjakan tentu kita berharap menguntungkan bagi kita. Tapi tentu tidak sama bagi orang lain

Gambar 3. Data komentar postingan terkait pembangunan tok Pemalang-Batang.

#### 3.2 Pre-processing

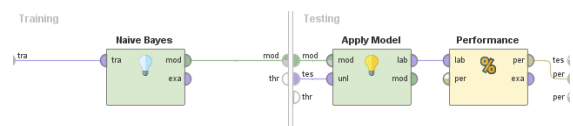
Proses pre-processing untuk tahapan case folding dilakukan dengan menggunakan microsoft excel, case folding dilakukan untuk mengubah teks komentar menjadi huruf kapital ataupun huruf kecil. Dalam penelitian ini digunakan teks komentar diubah kedalam huruf kecil semua.



Gambar 4. Tahapan Pre-processing

#### 3.3 Validasi

Tahap validasi atau pengujian dilakukan dengan menggunakan k fold cross validation. Cross validation akan membagi data menjadi k bagian untuk data training dan data testing. Proses ini akan diulangi sebanyak k yang telah ditentukan. Pada penelitian ini dilakukan beberapa eksperimen terhadap nilai k untuk melihat hasil yang terbaik dalam proses klasifikasi mengolah opini terkait pembangunan jalan tol Pemalang-Batang. Nilai k ditentukan dari k = 2 sampai k = 10.



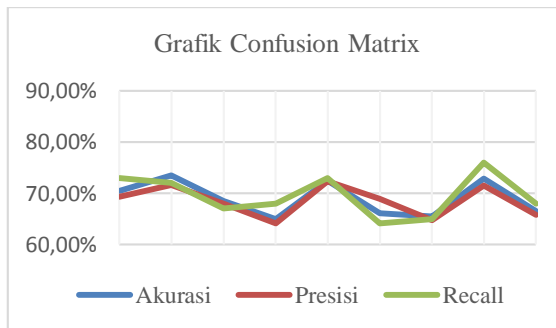
Gambar 5. Tahapan validasi model yang digunakan pada penelitian

### 3.4 Analisa Hasil Pengujian

Setelah dilakukan beberapa kali percobaan, diperoleh hasil confusion matrix yang meliputi akurasi, presisi dan recall dari model yang digunakan.

Tabel 2 Hasil Confusion Matrix

Fold	Confusion Matrix		
	Akurasi	Presisi	Recall
2	70.50%	69.29%	73.00%
3	<b>73.54%</b>	71.65%	72.04%
4	68.50%	67.97%	67.00%
5	65.00%	64.13%	68.00%
6	72.47%	<b>72.34%</b>	72.98%
7	66.06%	68.95%	64.08%
8	65.50%	64.80%	64.98%
9	72.92%	71.53%	<b>76.01%</b>
10	66.50%	65.75%	68.00%



Gambar 6. Grafik Confusion Matrix

Dari hasil confusion matrix pada tabel 2 dan gambar 6 nilai akurasi yang tertinggi adalah pada fold ke 3 dengan 73,54%, nilai presisi yang tertinggi diperoleh pada fold ke 6 dengan nilai 72,34% sedangkan untuk nilai recall tertinggi berada pada fold ke 9 dengan nilai 76,01%. Namun untuk hasil confusion matrix dengan rata-rata nilai yang sama ada pada fold ke 6.

Tabel 3 Hasil Confusion Matrix Fold 6

	true positif	true negatif	class precision
<b>pred. positif</b>	72	27	72.73%
<b>pred. negatif</b>	28	73	72.28%
<b>class recall</b>	72.00%	73.00%	

### 3.5 Analisa Hasil Opini

Tabel 4 Hasil Sentimen Positif

Kata	Bobot
baik	0.093833
alhamdulillah	0.093765
bagus	0.083312
indah	0.075983
waktunya	0.075983
lagi	0.075630
selamanya	0.074850
maju	0.074837
syukur	0.074651
tidak	0.074111

Tabel 5 Hasil Sentimen Negatif

Kata	Bobot
mahal	0.113538
ngutang	0.101103
desaku	0.095510
gratis	0.087385
bayar	0.085722
grosir	0.085496
sepi	0.084868
jalannya	0.080394
siapa	0.079087
jamannya	0.077454

Pada tabel 4 dan tabel 5 diatas menunjukkan distribusi kata-kata yang sering muncul pada dokumen komentar, terkait opini positif maupun opini negatif dari proyek pembangunan jalan tol Pemalang-Batang menggunakan algoritma naïve bayes.

## 4. KESIMPULAN DAN SARAN

### 4.1 Kesimpulan

Dari hasil penelitian yang dilakukan dengan menggunakan rapidminer 7.6 terkait analisis sentimen terhadap pembangunan jalan tol Pemalang-Batang di sosial media facebook dengan metode naïve bayes diperoleh nilai akurasi tertinggi mencapai 73,54% dengan nilai fold 3.

## 4.2 Saran

Saran yang perlu dipertimbangkan untuk penelitian berikutnya bisa menambahkan part of speech tagging atau pos tagging. Tahap ini digunakan untuk proses penentuan label dari kalimat dan juga untuk mengkategorikan gramatikan, seperti kata benda, kata kerja, kata sifat dan sebagainya.

## 5. REFERENSI

- Buntoro, Ghulam Asrofi. 2016. "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter." *Integer Journal Maret 1* (1): 32–41.  
[https://www.researchgate.net/profile/Ghulam\\_Buntoro/publication/316617194\\_Analisis\\_Sentimen\\_Calon\\_Gubernur\\_DKI\\_Jakarta\\_2017\\_Di\\_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf](https://www.researchgate.net/profile/Ghulam_Buntoro/publication/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf).
- Dang, Shilpa, and Peerzada Hamid Ahmad. 2014. "Text Mining : Techniques and Its Application." *IJETI International Journal of Engineering & Technology Innovations 1* (4): 22–25.  
[www.IJETI.com%0Awww.ijeti.com](http://www.IJETI.com%0Awww.ijeti.com).
- Gusriani, Syahmia, Kartina Diah Kusuma Wardhani, and Muhammad Ihsan Zul. 2016. "Analisis Sentimen Terhadap Toko Online Di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes." *4th Applied Business and Engineering Conference 1* (1): 1–7.
- Hall, Tracy, Sarah Beecham, David Bowes, David Gray, and Steve Counsell. 2011. "A Systematic Review of Fault Prediction Performance in Software Engineering." *IEEE Transactions on Software Engineering 38* (6): 1276–1304.  
<https://doi.org/10.1109/TSE.2011.103>.
- Han, Jiawei, Micheline Kamber, and Jian Pei. 2012. *Data Mining: Concepts and Techniques*. [https://doi.org/10.1002/1521-3773\(20010316\)40:6<9823::AID-ANIE9823>3.3.CO;2-C](https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C).
- Knight-mccord, Jasmine, Dylan Cleary, Nastassja Grant, Success Jumbo, Tiffany Lacey, Torri Livingston, Sky Robinson, Renardo Smith, and Richard Emanuel. 2016. "What Social Media Sites Do College Students Use Most?" *Journal of Undergraduate Ethnic Minority Psychology*, no. October 2015: 21–26.  
[http://www.juempyschology.com/wp-content/uploads/2016/05/Knight-McCord-et-al\\_JUEMP\\_2016.pdf](http://www.juempyschology.com/wp-content/uploads/2016/05/Knight-McCord-et-al_JUEMP_2016.pdf).
- Lidya, Syahfitri Kartika, Opim Salim Sitompul, and Syahril Efendi. 2015. "Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine ( Svm )." *Seminar Nasional Teknologi Dan Komunikasi 2015 2015* (Sentika): 1–8.  
<https://doi.org/10.1016/j.eswa.2013.08.047>.
- Mudry, Albert, and Anders Tjellström. 2011. "Historical Background of Bone Conduction Hearing Devices and Bone Conduction Hearing Aids." *Advances in Oto-Rhino-Laryngology 71*: 1–9.  
<https://doi.org/10.1159/000323569>.
- Novantirani, Anita, Mira Kania Sabariah, and Veronikha Effendy. 2015. "Analisis Sentimen Pada Twitter Untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota Dengan Metode Support Vector Machine" 2 (1): 1–7.
- Singh, Sarbjeet. 2010. *Artificial Intelligence*. Vol. 6.
- Wilson, Christo, Alessandra Sala, Krishna P. N. Puttaswamy, and Ben Y. Zhao. 2012. "Beyond Social Graphs." *ACM Transactions on the Web 6* (4): 1–31.  
<https://doi.org/10.1145/2382616.2382620>.
- Zaidi, Na, and J Cerquides. 2013. "Alleviating Naive Bayes Attribute Independence Assumption by Attribute Weighting." *The Journal of Machine ... 14*: 1947–88.  
<http://dl.acm.org/citation.cfm?id=2567725>.
- Zhang, Lei, and Bing Liu. 2016. "Sentiment Analysis and Opinion Mining." *Encyclopedia of Machine Learning and Data Mining*, no. May: 1–10.  
[https://doi.org/10.1007/978-1-4899-7502-7\\_907-1](https://doi.org/10.1007/978-1-4899-7502-7_907-1).
- Jati, Yusuf Waluyo. 2016. <http://industri.bisnis.com>. 06 17. Accessed 04 19, 2018.

<http://industri.bisnis.com/read/20160617/45/558946/tol-pemalang-batang-dan-batang-semarang-mulai-dibangun>.