



**ANALISIS SENTIMEN PADA *TWITTER* GOJEK DENGAN METODE
NAÏVE BAYES CLASSIFIER MENGGUNAKAN VISUALISASI *LATENT*
*DIRICHLET ALLOCATION***

JURNAL ILMIAH

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Statistika

Oleh :

**Shendy Sandra Muel
B2A016033**

**PROGRAM STUDI S1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS MUHAMMADIYAH SEMARANG
2020**

SURAT PERNYATAAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini, saya :

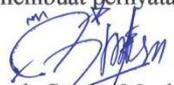
Nama : Shendy Sandra Muel
NIM : B2A016033
Fakultas/Jurusan : FMIPA/S1 - Statistika
Jenis Penelitian : Skripsi
Judul : Analisis Sentimen Pada *Twitter* Gojek dengan Metode *Naïve Bayes Classifier* Menggunakan Visualisasi *Latent Dirichlet Allocation*
Email : shendysandram@gmail.com

Dengan ini menyatakan bahwa saya menyetujui untuk :

1. Memberikan hak bebas royalti kepada Perpustakaan UNIMUS atas penulisan karya ilmiah saya, demi pengembangan ilmu pengetahuan
2. Memberikan hak menyimpan, mengalih mediakan/mengasih formatkan, mengelola dalam bentuk pangkalan data (database), mendistribusikannya, serta menampilkannya dalam bentuk *softcopy* untuk kepentingan akademis kepada Perpustakaan UNIMUS, tanpa perlu izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta
3. Bersedia dan menjamin untuk menanggung secara pribadi tanpa melibatkan pihak Perpustakaan UNIMUS dari semua bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini.

Demikian pernyataan ini saya buat dengan sesungguhnya dan semoga dapat digunakan sebagaimana mestinya.

Semarang, 06 Mei 2020
Yang membuat pernyataan,


Shendy Sandra Muel
(B2A016033)

ANALISIS SENTIMEN PADA *TWITTER* GOJEK DENGAN METODE *NAIVE BAYES CLASSIFIER* MENGGUNAKAN VISUALISASI *LATENT DIRICHLET ALLOCATION*

Shendy Sandra Muel ⁽¹⁾, Tiani Wahyu Utami ⁽²⁾ dan Rochdi Wasono ⁽³⁾

^(1,2,3) Program Studi Statistika Universitas Muhammadiyah Semarang

Email : shendysandram@gmail.com

Article history	Abstract
Submission :	Gojek is an online based transportation that already has a mobile application, and has an account on social media, thus providing a gojek user to provide an opinion. The number of opinions on social media requires a method that can categorize these opinions automatically, including positive or negative. Sentiment analysis is a process in changing contextual texts, this determines a text has a positive or negative meaning. Naïve Bayes Classifier is a method of extracting text that is used and its processing time is fast and easy to implement with a fairly easy structure and a high level of effectiveness. Latent Dirichlet Allocation is a statistical method used to find out trending topics in a document. The purpose of this research is to get the results of the classification of Twitter subject sentiments and trends of topics that are often discussed by Twitter users on the subject. The results obtained by the majority of Twitter users gave a positive response to the subject with the results obtained using the Naïve Bayes Classifier method of 78.55% and the topic trends obtained using the Latent Dirichlet Allocation related to words about features. Evidenced by the appearance of the words "gojek", "voucher", "need", "food", and "goride". From this it can be seen that Twitter social media users use food and food vouchers from Gojek.
Revised :	
Accepted :	

Keyword:
Sentiment Analysis,
Gojek, Latent
Dirichlet Allocation,
Naïve Bayes
Classifier, Topic
Modeling.

PENDAHULUAN

Transportasi merupakan hal yang erat hubungannya dengan masyarakat. Transportasi terdapat berbagai macam jenisnya, baik roda dua maupun roda empat. Transportasi umum juga masih digunakan masyarakat sebagai sarana untuk menuju suatu tempat walaupun untuk memiliki kendaraan pribadi di zaman sekarang sangatlah mudah (Nugroho, et al 2016). Berkembangnya teknologi saat ini memunculkan aplikasi transportasi berbasis *online* salah satunya yaitu aplikasi gojek. Gojek merupakan transformasi dari ojek konvensional yang

bertempat di suatu pangkalan untuk menunggu pelanggan, atau bisa disebut ojek pangkalan. Kemudahan dalam pemesannya dengan menggunakan teknologi *smartphone* membuat gojek semakin banyak diminati, hal ini dapat menjadi solusi transportasi di Indonesia terutama di kota besar yang sering sekali mengalami kemacetan.

Gojek merupakan sebuah perusahaan teknologi asal Indonesia yang menyediakan jasa transportasi ojek, saat ini aplikasi gojek telah diunduh oleh 170 juta pengguna di Indonesia dan Asia Tenggara (cnbcindonesia.com). Gojek Indonesia merupakan satu-satunya perusahaan

dari Asia Tenggara yang masuk dalam daftar 56 perusahaan yang mengubah dunia atau disebut dengan daftar “*Change the World*”, dimana gojek menempati peringkat ke-17, bersama sejumlah perusahaan-perusahaan kelas dunia lainnya (gojek.com).

Opini yang dimuat di media sosial jumlahnya terlalu banyak. Oleh sebab itulah, diperlukan sebuah metode atau teknik khusus yang mampu mengkategorikan *review-review* tersebut, apakah termasuk positif atau negatif (Potdar, et al 2016 dalam Khotimah 2019), serta topik apa saja yang sering dikeluhkan oleh pengguna media sosial mengenai layanan pada perusahaan Gojek Indonesia.

Analisis sentimen adalah suatu proses dalam mendeteksi teks secara kontekstual, ini menentukan suatu teks memiliki makna positif dan negatif. Teknik ini digunakan di banyak bidang seperti *e-commerce*, perawatan kesehatan, hiburan dan politik. Penelitian ini akan menggunakan metode *Naïve Bayes Classifier* dengan visualisasi *Latent Dirichlet Allocation*. Penggunaan *Naïve Bayes Classifier* dalam proses klasifikasi memiliki kelebihan dalam kemudahan penggunaannya serta waktu pemrosesannya yang cepat, mudah diimplementasikan dengan struktur yang cukup sederhana dan tingkat efektifitas yang tinggi. *Latent Dirichlet Allocation* merupakan sebuah metode yang tergolong sebagai salah satu bentuk dari *text mining* untuk menemukan pola tertentu pada sebuah dokumen. *Latent Dirichlet Allocation* membuat dokumen menghasilkan beberapa macam topik yang berbeda, sehingga tidak secara spesifik mengelompokkan dokumen kedalam sebuah topik tertentu. Hasil dari analisis pemodelan topik menggunakan *Latent Dirichlet Allocation* akan ditampilkan dengan sebuah visualisasi berupa *dashboard* pada media sosial yang dapat menampilkan *trend* dari topik-topik yang sering dibahas, sebagai suatu sistem yang dapat membantu pengambilan keputusan. Oleh karena itu, pada penelitian ini penulis akan melakukan Analisis Sentimen Pada *Twitter* Gojek dengan Metode *Naïve Bayes Classifier* Menggunakan Visualisasi *Latent Dirichlet Allocation*.

Beberapa penelitian mengenai Analisis Sentimen Pada *Twitter* Gojek dengan Metode *Naïve Bayes Classifier* Menggunakan Visualisasi *Latent Dirichlet Allocation* telah dilakukan oleh Agustina, A (2017) dengan judul “Analisis dan Visualisasi Suara Pelanggan Pada Pusat Layanan Pelanggan dengan Pemodelan Topik Menggunakan *Latent Dirichlet Allocation*

(LDA) Studi Kasus: PT. Petrokimia Gresik” dengan hasil pengujian yang didapatkan yaitu pengujian dokumen terhadap topik menghasilkan ketepatan sebesar 83.7% yaitu 190 dokumen dari 227 dokumen.

Penelitian yang dilakukan oleh Hadi et.al (2017) dengan judul penelitian : “*Text Mining* Pada Media Sosial *Twitter* Studi Kasus: Masa Tenang Pilkada DKI 2017 Putaran 2” dengan hasil yang didapatkan akurasi rata-rata sebesar 80%.

Penelitian yang dilakukan oleh Syakuro (2017) dengan judul penelitian : “Analisis Sentimen Masyarakat Terhadap *E-Commerce* Pada Media Sosial Menggunakan Metode *Naïve Bayes Classifier* (NBC) dengan Seleksi Fitur *Information Gain* (IG)” dengan hasil yang didapatkan akurasi sebesar 88.8%.

Penelitian yang dilakukan oleh Utami (2017) dengan judul penelitian : “Analisis Topik Data Media Sosial *Twitter* Menggunakan Model Topik *Latent Dirichlet Allocation*” dengan hasil yang didapatkan pemodelan topik LDA pada kelima lokasi *tweet* di Kota Bogor dan rentang waktu tertentu berhasil membentuk topik dengan informasi atau deskripsi topik untuk setiap lokasi *tweet*.

TINJAUAN PUSTAKA

Transportasi Online

Perkembangan teknologi secara pesat dalam bidang telekomunikasi di seluruh dunia menyebabkan setiap negara harus mampu bersaing dengan pemanfaatan teknologi serta mengaplikasikannya di dalam beraktivitas. Dalam bidang transportasi, beralihnya jasa transportasi konvensional ke jasa transportasi online merupakan bagian dari kemajuan teknologi. Diciptakannya teknologi bertujuan untuk memudahkan berbagai aktifitas manusia sehari-hari. Dengan kehadiran dari transportasi online menjawab kebutuhan masyarakat berkaitan dengan angkutan umum (Anwar 2017).

Twitter

Twitter merupakan salah satu situs jejaring sosial yang masih populer hingga saat ini. *Twitter* didirikan oleh Jack Dorsey pada bulan Maret 2006. *Twitter* membatasi kata yang akan di post sebanyak 140 karakter. Namun, tidak hanya tulisan yang dapat diunggah pada jejaring sosial tersebut, *twitter* juga bisa mengunggah foto, video, url, dan lain-lain.

$$P(C|E) = \frac{P(E|C)P(C)}{P(E)}$$

Keterangan:

$P(C|E)$: Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis C terjadi jika diberikan bukti (*evidence*) E terjadi.

$P(E|C)$: Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis C.

$P(C)$: Probabilitas awal (priori) hipotesis C terjadi tanpa memandang bukti apapun.

$P(E)$: Probabilitas awal (priori) hipotesis E terjadi tanpa memandang hipotesis bukti yang lain.

Metode *Naïve Bayes Classifier* merupakan penyederhanaan dari algoritma *Naïve Bayes* dan cocok digunakan dalam pengklasifikasian teks atau dokumen.

Ukuran Evaluasi Model Klasifikasi

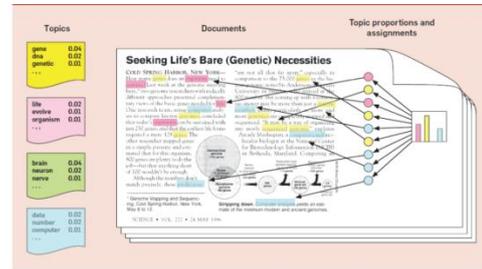
Evaluasi pada suatu klasifikasi pada umumnya dilakukan dengan menggunakan sebuah himpunan data yang diuji, tidak digunakan dalam pelatihan klasifikasi tersebut. Pada tahap ini terdapat sejumlah ukuran yang dapat digunakan untuk menilai kembali atau mengevaluasi model klasifikasi, yaitu *accuracy* atau tingkat pengenalan, tingkat kesalahan atau kekeliruan klasifikasi, *recall* atau *sensitivity* atau *true positif*, *specificity* atau *true negative* dan *precision*.

Pada proses evaluasi klasifikasi terdapat empat kemungkinan yang terjadi yaitu proses pengklasifikasian pada suatu baris data. Hasil klasifikasi biner pada suatu dataset yang dipresentasikan dalam bentuk matriks 2x2 yaitu dinamakan *confusion matrix*.

Topic Modelling

Topic Modelling atau pemodelan topik merupakan kumpulan algoritma yang digunakan untuk menemukan struktur tersembunyi dari tema yang terdapat dalam setiap dokumen. Algoritma ini, dapat digunakan untuk pengembangan dalam pencarian, ataupun meringkas teks yang ada dalam dokumen (Blei 2003 dalam Agustina 2017). *Topic Modelling* dapat mengatur kumpulan kata berdasarkan tema yang ditemukan. Selain itu, *topic modelling* juga dapat diaplikasikan untuk berbagai jenis data,

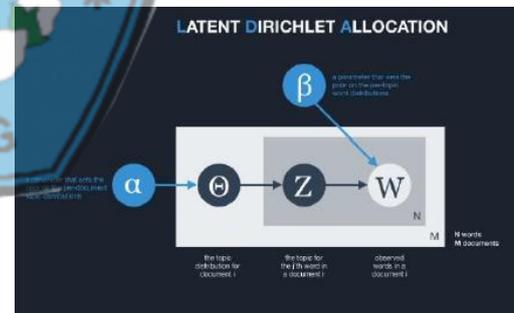
seperti saat ini dilakukan adalah untuk mencari pola pada data genetik, gambar, ataupun pada sosial media. Konsep *topic modelling* menurut Blei, ditunjukkan pada Gambar 2 :



Konsep *Topic Modelling*

Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) adalah metode statistika yang digunakan sebagai model untuk menganalisis suatu dokumen. LDA berusaha untuk melihat dokumen dengan cara mundur untuk menemukan satu set topik yang mungkin telah dikoleksi. LDA mempresentasikan dokumen dengan berbagai topik yang dibuat berdasarkan probabilitas tertentu (Chen 2011 dalam Agustina 2017). Probabilitas topik, merepresentasikan kejelasan dari suatu dokumen. Menurut Blei bahwa dalam sebuah dokumen terdapat beberapa komponen, yang ditunjukkan pada Gambar 3 :



Visualisasi Metode LDA

Ide dasar dari LDA adalah bahwa dalam dokumen, merepresentasikan campuran topik secara acak, dimana setiap topik digolongkan berdasarkan distribusi antar kata. LDA tidak hanya digunakan untuk melakukan pendeteksian topik saja, namun LDA juga digunakan sebagai salah satu *tools* untuk berbagai penelitian seperti pada konten percakapan. Sebagai salah satu contoh dari Blei, distribusi topik yang ditampilkan dengan kumpulan kata-kata pada dokumen ditunjukkan pada gambar 3.

METODE PENELITIAN

Sumber Data

Data yang digunakan adalah data sekunder. Data diambil dengan menggunakan proses *Scrapping* pada *twitter* dengan hastag Gojek. Populasi pada penelitian ini adalah seluruh *tweets* pengguna *twitter* mengenai Gojek. Sampel yang digunakan adalah data diambil sebanyak 5000 *tweets* pada Januari tahun 2020. Teknik pengumpulan data yang berasal dari pesan teks atau *tweet* dari timeline *twitter* diperoleh dari API (*Application Programming Interface*).

Variabel Penelitian dan Struktur Data

Data yang didapatkan dari hasil *crawling* data menggunakan API *twitter* dibagi menjadi data *training* dan data *testing* menggunakan *40-fold cross validation* dengan perbandingan 60% : 40%. Untuk struktur data yang digunakan dalam penelitian ini setelah dilakukan *preprocessing* pada data teks *tweet* dari variabel prediktor yaitu kata dasar setiap *tweet* dan variabel respon yaitu klasifikasi sentiment *tweet* (positif dan negatif). Berikut struktur data penelitian pada tabel 1 sebelum *preprocessing*.

Struktur Data Penelitian

No	Tweet	Sentimen
1	pesen makan juga akhirnya untung ada gojek	Positif
2	apasi voucher gojek tuh bingung ngitung potongannya gmn gajelas	Negatif

Langkah-langkah Penelitian

Langkah analisis data yang digunakan dalam penelitian ini sebagai berikut :

1. Mengambil data *tweet* dengan menggunakan *Twitter API*
 - a. Memasukkan *keyword* Gojek
 - b. Menyimpan hasil *crawling* ke *database*

2. *Text Prerocessing* yang terdiri dari proses *spelling normalization*, *case folding*, *tokenizing*, *filtering*, dan *stemming*.
3. Lalu dilakukan seleksi dengan TF-IDF yaitu mengubah teks menjadi vektor dan pembobotan kata.
4. Melakukan evaluasi klasifikasi *K-fold Cross Validation*.
5. Klasifikasi data menggunakan metode *Naive Bayes Classifier* untuk menentukan hasil akurasi.
6. Visualisasi menggunakan *Latend Dirichlet Allocation* untuk menemukan *trend* topik yang sering dibahas dalam teks.
7. Interpretasi dan mengambil kesimpulan.

HASIL PENELITIAN DAN PEMBAHASAN

Tahapan *Text Preprocessing*

Pada tahapan *text preprocessing* digunakan untuk melakukan seleksi data untuk mengubah data teks yang tidak terstruktur atau sembarang menjadi data yang terstruktur. Berikut hasil dari tahapan *text preprocessing* pada tabel 2 sebagai berikut :

Hasil *Text Preprocessing*

Hasil Case Folding dan Tokenisasi	
kl isi gopay dr mbanking tapi gak masuk, itu sy komplain ke bank nya apa ke gojek ya?	kalau isi gopay dari mobile banking tetapi tidak masuk itu saya komplain ke bank nya apa ke gojek ya
Hasil Stemming dan Filtering	
kalau isi gopay dari mobile banking tetapi tidak masuk itu saya komplain ke bank nya apa ke gojek ya	kalau isi gopay dari mobile banking tetapi tidak masuk saya komplain bank apa gojek
Hasil Data Bersih	
kalau isi gopay dari mobile banking tetapi tidak masuk saya komplain bank apa gojek	kalau isi gopay mobile banking tetapi tidak masuk komplain bank apa gojek

Pelabelan Data

Pelabelan pada kelas sentimen menggunakan kategori positif dan kategori negatif, adapun hasil dari pelabelan untuk gojek dapat ditunjukkan tabel 3 sebagai berikut :

Jumlah Pelabelan Kelas Sentimen	
Sentimen	Tweet Gojek
Positif	2684
Negatif	2316

Berdasarkan rekap jumlah pelabelan kelas sentimen didapatkan bahwa sentiment positif untuk review gojek sebanyak 2684 tweets sedangkan untuk sentiment negatif gojek sebanyak 2316 tweets. Tweets diklasifikasikan sebagai sentiment positif apabila mengandung pernyataan positif seperti pujian, ungkapan terima kasih, atau testimony positif tentang gojek. Hasil dari pelabelan dapat dilihat pada tabel 4 dibawah ini.

Hasil Pelabelan Kelas Sentimen		
Sentimen	Skor	Text
Negatif	-3	apasi voucher gojek tuh bingung ngitung potongannya gmn gajelas
Negatif	-1	gojek demi allah gue gabutuh voucher makanan gue butuhnya goride
Positif	1	pesen makan juga akhirnya untung ada gojek
Positif	0	kalo grab masih sering dapett sih cuma sbg mahasiswa ngirit pake gojek lebih enak karena bisa paylater

Klasifikasi

Pada tahap ini dilakukan evaluasi pada klasifikasi menggunakan metode *Naïve Bayes Classifier*. Hasil pertama diperoleh persamaan pada *confusion matrix* klasifikasi pada setiap kelas. Berikut hasil yang didapatkan :

Confusion Matrix Klasifikasi Gojek

	Kenyataan Positif	Kenyataan Negatif	Jumlah
Prediksi	1125	394	1519

Positif			
Prediksi	35	446	481
Negatif			
Jumlah	1160	840	2000

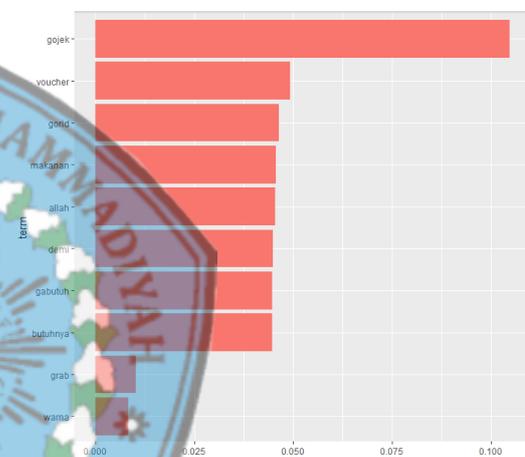
Berdasarkan tabel *Confusion matrix* diatas prediksi benar sebanyak 1571 dan prediksi salah sebanyak 429.

Hasil Klasifikasi dengan *Naïve Bayes Classifier*

Hasil	Gojek
Accuracy	78.55%

Hasil klasifikasi dengan metode *Naïve Bayes Classifier* untuk Gojek sebesar 78.55%.

Latent Dirichlet Allocation



Hasil *Latent Dirichlet Allocation*

Visualisasi yang dihasilkan menggunakan *Latent Dirichlet Allocation* pada bulan Januari 2020 didapatkan yaitu terdapat trend topik mengenai kata-kata tentang fitur. Dibuktikan dengan kemunculan kata “gojek”, “voucher”, “butuhnya”, “makanan”, dan “gorid”. Dari sini dapat diketahui bahwa pengguna media sosial *twitter* membutuhkan voucher makanan dan goride dari Gojek.

KESIMPULAN DAN SARAN

Kesimpulan

1. Hasil pengklasifikasian sentimen pada *twitter* gojek menggunakan metode *Naïve Bayes Classifier* didapatkan jumlah sentimen positif sebanyak 2684 dan sentimen negatif

sebanyak 2316 yang digambarkan sebagai berikut :

- a. Sentimen positif: “pesen makan juga akhirnya untung ada gojek”.
- b. Sentimen negatif: “apasi voucher gojek tuh bingung ngitung potongannya gmn gajelas”.

Pembagian data 40 *Fold Cross Validation* diperoleh tingkat akurasi sebesar 78.55%. Hal ini menunjukkan kemampuan prediksi dengan menggunakan *Naïve Bayes Classifier* baik sehingga dapat dijadikan sebagai metode rujukan untuk mengolah data *text*.

2. Visualisasi yang dihasilkan menggunakan *Latent Dirichlet Allocation* pada bulan Januari 2020 didapatkan yaitu terdapat *trend* topik mengenai kata-kata tentang fitur. Dibuktikan dengan kemunculan kata “gojek”, “voucher”, “butuhnya”, “makanan”, dan “gorid”. Dari sini dapat diketahui bahwa pengguna media sosial *twitter* membutuhkan voucher makanan dan goride dari Gojek.

Saran

Berdasarkan hasil analisis dan kesimpulan, dapat diberikan saran sebagai berikut :

1. Bagi pihak Gojek, hasil dari informasi dari ulasan-ulasan yang telah diberikan oleh pengguna dapat dijadikan sebagai bahan evaluasi dalam peningkatan kepuasan pengguna dan memberikan pelayanan semaksimal mungkin, serta untuk pengembangan pembaharuan aplikasi serta fitur-fitur selanjutnya.
2. Untuk penelitian selanjutnya sebaiknya dapat menggunakan sistem pelabelan yang memiliki tingkatan lebih tinggi, yakni mampu mendeteksi sentiment pada frasa dan kalimat.
3. Untuk penelitian selanjutnya dengan kasus sentimen pada Gojek sebaiknya menggunakan metode lain selain *Naïve Bayes Classifier* untuk mendapatkan akurasi yang lebih baik.
4. Menerapkan *Latent Dirichlet Allocation* untuk beragam data dan studi kasus menarik lainnya.

DAFTAR PUSTAKA

Agustina, A.2017.Analisis dan Visualisasi Suara Pelanggan Pada Pusat Layanan Pelanggan dengan Pemodelan Topik

Menggunakan *Latent Dirichlet Allocation* (LDA) Studi Kasus: PT. Petrokimia Gresik. Skripsi.Surabaya: Institut Teknologi Sepuluh November.

Anwar, A, A. 2017. Online VS Konvensional: Keunggulan dan Konflik Antar Moda Transportasi di Kota Makassar.

Fatimah, A.2018 . Penerapan Metode DBSCAN Dalam Memperbaiki Kinerja *K-Means* Untuk Penggerombolan Data *Tweet*. Skripsi.Bogor: Institut Pertanian Bogor.

Ghifari, A. 2018. Analisis Popularitas IPB Berdasarkan Data Twitter Menggunakan Awan Kata, Geovisualisasi, Dan Pemodelan Klasifikasi. Skripsi.Bogor: Institut Pertanian Bogor.

Hadi, A. F., Bagus. D., dan Hasan, M.2017. Text Mining Pada Media Sosial Twitter Studi Kasus : Masa Tenang Pilkada DKI 2017 Putaran 2, Semin. Nas. Mat. dan Apl. 21 Oktober 2017 Surabaya: Universitas Airlangga.

Ibrahim, N., dan Wibowo, A. 2014. *Support Vector Regression with Missing Data Treatment Based Variables Selection for Water Level Prediction of Galas River in Kelantan Malaysia*. Johor Bahru: Universitas Teknologi Malaysia.

Kamal, A.F dan Widjajanto, B.2017.*Text Mining Untuk Analisa Sentiment Ekspedisi Jasa Pengiriman Barang Menggunakan Metode Naive Bayes Pada Aplikasi J & T Express*.

Khotimah, N. 2019. Analisis Sentimen Terhadap *Review E-Commerce* dengan Metode *Stochastic Gradient Descent*. Skripsi.Semarang: Universitas Muhammadiyah Semarang)

Kurniawan, T. 2017. Implementasi *Text Mining* Pada Analisis Sentimen Pengguna Twitter Terhadap Media *Mainstream* Menggunakan *Naïve Bayes Classifier* Dan *Support Vector Machine*. Tugas Akhir. Surabaya: Institut Teknologi Sepuluh November.

Kurniawan, W. 2018. Sistem Monitoring Percakapan pada Toko Online Menggunakan Metode *Latent Dirichlet Allocation* (LDA) Studi Kasus: Toko Online “BerryBenka.com”. Tugas Akhir. Yogyakarta: Universitas Islam Indonesia Yogyakarta.

Liu, Y., Loh, H. T., dan Sun, A. 2009. Imbalanced text classification: A term weighting approach. *Expert Systems with Applications*, 36(1), 690–701.

- <https://doi.org/10.1016/j.eswa.2007.10.042>
- Nugroho, D. G., Chrisnanto, Y. H., dan Wahana, A. 2016. Analisis Sentimen Pada Jasa Ojek *Online* Menggunakan Metode *Naïve Bayes*
- Nurhuda, F., Sihwi, S. W., dan Doewes, A. 2013. Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode *Naïve Bayes Classifier*.
- Pradana, M. 2015. Klasifikasi Bisnis *E-Commerce* Di Indonesia. *Modus* Vol.27 (2): 163-174, 2015
- Praptiwi, D. Y. 2018. Analisis Sentimen *Online Review* Pengguna *E-Commerce* Menggunakan Metode *Support Vector Machine* dan *Maximum Entropy* (Studi Kasus: Review Bukalapak pada Google Play). Yogyakarta: Universitas Islam Indonesia.
- Prihatini, P. M., Suryawan, I. K., dan Mandia, I. N. 2017. Metode Latent Dirichlet Allocation Untuk Ekstraksi Topik Dokumen. *J. Log.*, vol. 17, no. 3, pp. 154–158
- Putra, A. P. 2017. Analisis Sentimen Data Twitter Menggunakan *Naive Bayes* dengan *Negation Handling* Pada Data Twitter Bahasa Indonesia. *Skripsi*. Bogor: Institut Pertanian Bogor.
- Putranti, N. D., dan Winarko, E. 2014. Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan *Maximum Entropy* dan *Support Vector Machine*. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 8(1), 91–100. <https://doi.org/10.22146/ijccs.3499>
- Putri, D.U.K. 2016. *Implementasi Inferensi Fuzzy Mamdani Untuk Keperluan Sistem Rekomendasi Berita Berbasis Konten*. Skripsi. Program Studi Ilmu Komputer FMIPA UGM Yogyakarta
- Saputra, F. T. 2018. Analisis Sentimen Pada Twitter Untuk Memprediksi Hasil Pilkada Menggunakan *Rule-Based Classification* Berbasis Sentimen Leksikon Indonesia.
- Saraswati, N.S. 2011. *Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis*. Skripsi. Program Studi Teknologi Informasi Fakultas Teknik UGM Yogyakarta.
- Syakuro, A. 2017. Analisis Sentimen Masyarakat Terhadap *E-Commerce* Pada Media Sosial Menggunakan Metode *Naïve Bayes Classifier* (NBC) dengan Seleksi Fitur *Information Gain* (IG) Halaman Judul Skripsi Oleh: Abdan Syakuro. *Skripsi*. Malang: Universitas Islam Negeri Maulana Malik Ibrahim.
- Utami, K. P. 2017. Analisis Topik Data Media Sosial Twitter Menggunakan Model Topik *Latent Dirichlet Allocation*. *Skripsi*. Bogor: Institut Pertanian Bogor.
- Zafikri, A. 2008. *Implementasi Metode Term Frequency Inverse Document Frequency (TF-IDF) Pada Sistem Temu Kembali Informasi*. Skripsi. Program Studi S-1 Ilmu Komputer FMIPA USU.
- Zaki, M. J., dan Meira, W. J. 2014. *Data Mining and Analysis: Fundamental Concepts and Algorithms*, 562. Retrieved from <https://books.google.com/books?hl=en&lr=&id=Gh9GAwAAQBAJ&pgis=1>