

Analisis Sentimen *Self-driving car* dengan Sentiment Confident Terbaik

Nurirwan Saputra
Teknik Informatika
Universitas PGRI Yogyakarta
Indonesia
nurirwan@upy.ac.id

Abstrak— *Self-driving car* merupakan inovasi dalam teknologi transportasi khususnya kendaraan roda empat yang merupakan bagian dari revolusi Industri 4.0. *Self-driving car*. *Self-driving car* mulai banyak bermunculan dari berbagai macam perusahaan, diantaranya yang cukup besar adalah perusahaan Google, Tesla, CMU, Toyota, Audi, Volvo, Nissan dan Mercedes Benz.

Terkait dengan mulai banyaknya perusahaan yang mengembangkan *Self-driving car*, banyak masyarakat yang memberikan pandangannya terkait dengan teknologi ini. Dari data yang didapat dari <https://data.world> terkait dengan *Self-driving car* memiliki lima class attribute, diantaranya sangat negative, agak negative, netral, agak positif, sangat positif. Dari data tersebut diambil hanya data yang memiliki nilai sentiment confident terbaik, yaitu satu. Metode klasifikasi yang digunakan adalah metode bayes, yaitu Naïve Bayes, Bayes Net, Naïve Bayes Multinomial, Naïve Bayes Multinomial Text dan Naïve Bayes Multinomial Updateable.

Hasil klasifikasi yang didapat dengan metode Naïve Bayes sebesar 67,92, Bayes Net sebesar 79,84, Naïve Bayes Multinomial sebesar 82,23, Naïve Bayes Multinomial Text sebesar 78,11, Naïve Bayes Multinomial Updateable 82,23, dan Naïve Bayes Updateable 67,92.

Kata kunci— *self-driving car*, *sentiment analysis*, *analysis sentiment*, *mobil otomatis*.

I. PENDAHULUAN

Mobil tanpa pengemudi atau dikenal dengan istilah *self-driving car* menjadi teknologi yang hangat dibicarakan dan dikembangkan akhir-akhir ini. Sistem yang ada pada *self-driving car* secara otomatis mempelajari representasi internal dari proses dan langkah-langkah seperti yang dilakukan oleh pengemudi mobil [1]. Dengan adanya *self-driving car*, kecelakaan yang terjadi karena kesalahan manusia di jalan raya dapat lebih diminimalkan, bahkan hingga sampai 90% [2]. Sehingga dengan adanya fakta ini, dapat lebih menguntungkan baik pengemudi maupun pengguna jalan lain.

Google telah melakukan eksperimen *self-driving car* dalam pengumpulan data Google StreetView (Gambar 1). Mobil yang telah menempuh perjalanan sepanjang 2.9 juta km hanya terdapat 13 kecelakaan minor [3]. Teknologi Big Data memegang peran sangat penting dalam perkembangan teknologi ini. Dengan berbagai sensor yang begitu banyak tertanam, data yang dikumpulkan per detiknya adalah 1 Gigabyte dan dalam setahun dapat mencapai 2 PetaByte. Itu baru 1 mobil saja, bayangkan jika seluruh mobil rencananya akan saling terhubung satu sama lain. Invent, sebuah perusahaan yang melakukan riset dalam teknologi inter-vehicular mengatakan, kendaraan -atau sistem kontrol mereka- akan melakukan sinkronisasi secara konstan dengan setiap lingkungan dan kondisi jalan. Kendaraan tersebut akan

mengetahui jika terdapat hambatan atau kemacetan, dan dapat mengubah rute secara real time [3].



Gambar 1. *Self-driving car* Google Street View [4]

Perusahaan Apple juga ikut mengembangkan self driving car ini tetapi tidak mengembangkan dari awal melainkan bekerjasama dengan BMW perusahaan besar yang bergerak di bidang otomotif. Dengan keikutsertaan evarini dalam mengembangkan self Driving Cars maka pandang sebelah mata dan pusat perhatian dunia. Project tingkat Apple dinamakan Titan. Selain itu Apple juga merekrut seorang pakar software otomatis motif di BlackBerry bernama Dodge dalam pengembangan software. BMW i3 diketahui sebagai mobil yang akan dijadikan pengembangan automous Project Titan [5].

Walaupun dengan banyaknya manfaat dari *self-driving car*, ada beberapa kontradiksi yang menyebar di masyarakat, diantaranya terkait dengan demografi dan penggunaan mobil, masyarakat rela untuk mengendarai mobil hanya belum siap untuk membelinya dikarenakan pengemudi masih tidak percaya dengan teknologi *self-driving car*. Masih banyak masyarakat yang memiliki kekhawatiran bahwa teknologi *self-driving car* tidak aman dan tidak canggih. Untuk itu perusahaan pengembang *self-driving car* harus meluncurkan teknologi ini secara bertahap untuk meningkatkan keamanan dan kepercayaan masyarakat [6].

Dikarenakan *self-driving car* akan menjadi teknologi yang akan menggantikan manual driving, maka perlu juga menampung aspirasi masyarakat terkait dengan teknologi ini. Penelitian ini membahas terkait analisis sentimen pandangan masyarakat terhadap teknologi baru ini, yaitu *self-driving car* dengan 5 class attribute yaitu sangat positif, agak positif, netral, agak negative, dan sangat negative. Selain itu, penelitian ini hanya menggunakan data dengan tingkat kepercayaan atau sentiment confidence nya tertinggi, yaitu dengan nilai 1.

II. TINJAUAN PUSTAKA

A. Data Mining

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, pembelajaran mesin untuk mengekstraksi dan mengidentifikasi informasi yang berguna dan pengetahuan terkait dari basis data besar. Penambangan data adalah proses menemukan pola baru dalam data dengan memfilter data dalam jumlah besar. Penambangan data menggunakan teknologi pengenalan pola yang mirip dengan teknik statistik dan teknik matematika. Pola yang ditemukan dapat memberikan informasi yang berguna untuk menghasilkan manfaat, keefektifan, dan efisiensi ekonomi [7]. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses Knowledge Discovery in Database (KDD) secara keseluruhan [8].

B. Analisis Sentimen

Analisis Sentimen merupakan cabang ilmu data mining yang masih ada kaitannya dengan text mining. Analisis Sentimen memahami opini masyarakat baik sentiment, perilaku maupun emosi [9] yang masih dalam bentuk text untuk dapat dipahami oleh komputer nilai positif maupun negative dari opini yang tertulis tersebut. Dalam cabang ilmu ini, biasanya terkait dengan opini masyarakat baik yang positif maupun yang negative. Analisis Sentimen sudah mulai trend di era saat ini, seperti contohnya untuk manajemen reputasi [10] maupun opini mahasiswa terhadap Universitasnya [11].

C. Naïve Bayes

Naïve Bayes merupakan salah satu algoritma dalam Data Mining atau Machine Learning yang menggunakan pendekatan probabilitas. Naïve Bayes banyak digunakan dikarenakan algoritma ini cukup sederhana, efisiensi, dan akurat [12]. Naïve Bayes menjadi salah satu dari 10 algoritma teratas dalam komunitas Data Mining ataupun Machine Learning [13] dan pembobotannya lebih menekankan fitur prediktif daripada yang kurang prediktif [14].

D. Naïve Bayes Updateable

Kelas yang digunakan untuk klasifikasi Naïve Bayes menggunakan estimasi kelas. ini adalah perbedaan dari algoritma Naïve Bayes Updateable dibanding Naïve Bayes yang umum. Klasifikasi ini menggunakan nilai presisi default yaitu 0,1 untuk atribut/data numerik ketika memanggil fungsi buildClassifier dengan instance pelatihan nol [15].

E. Bayesian Network

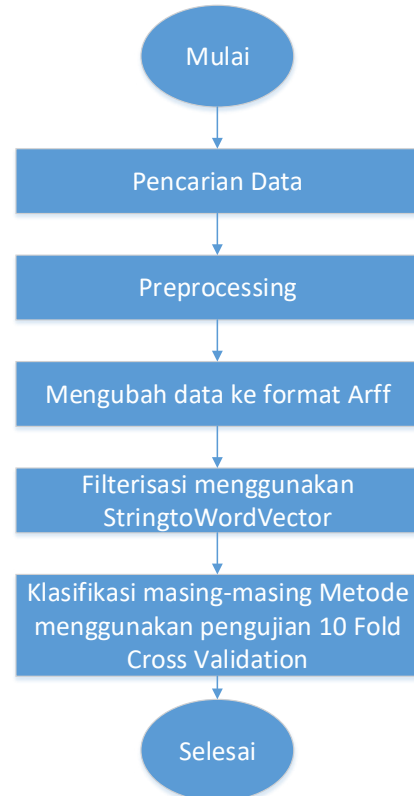
Bayesian Network adalah jenis model grafis probabilistik yang menggunakan inferensi Bayesian untuk perhitungan probabilitas. Bayesian Network bertujuan untuk memodelkan ketergantungan bersyarat dalam grafik. Melalui hubungan ini, seseorang dapat secara efisien melakukan inferensi pada variabel acak dalam grafik melalui penggunaan factor [16]. Bayesian Network diusulkan untuk berfungsi sebagai alat pendukung pengambilan keputusan untuk studi masa depan [17].

F. Naïve Bayes Multinomial

Naïve Bayes Multinomial merupakan metode dalam Data Mining yang pembelajarannya diawasi (Supervised Learning) [9]. Naïve Bayes Multinomial banyak digunakan untuk penelitian yang berbasis text, seperti filter pesan berupa spam [18].

III. METODOLOGI PENELITIAN

Langkah-langkah yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 2. Langkah-langkah dalam penelitian

A. Pencarian Data

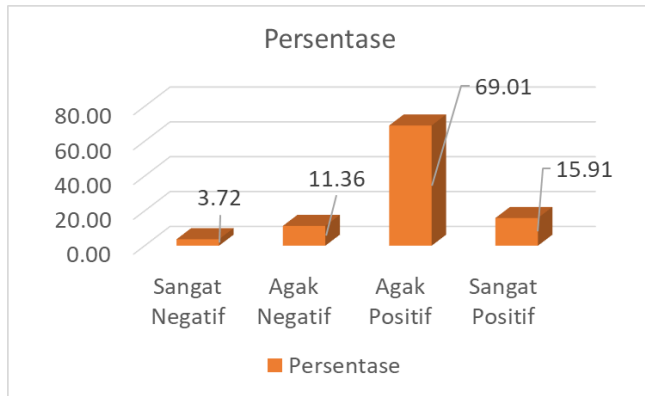
Data yang akan diolah pada penelitian ini merupakan data sekunder yang diambil dari bank data yang beralamat di <https://data.world/crowdfunder/sentiment-self-driving-cars> [19].

B. Preprocessing

Data yang sudah diambil ini dilakukan preprocessing terlebih dahulu diantaranya akan dilakukan eliminasi field yang tidak berkaitan dengan analisis sentiment seperti sentiment gold Reason, kemudian field our ID, field last Judgement dan selanjutnya field unit id dan menghapus semua data pada field United karena hanya berisi golden. Langkah selanjutnya adalah dari data yang sudah dilakukan eliminasi selanjutnya akan dieliminasi data yang tidak relevan jadi yang sentiment nya itu tidak relevan. Record data awal masih ada 6945 record dari sini agar akurasi dan tingkat confidence tinggi, maka akan dieliminasi lagi data hanya yang memiliki sentiment confidence nya yang nilainya 1, sehingga record saat ini berjumlah 2679 record. Mengambil data yang hanya sentiment confidence nya tinggi

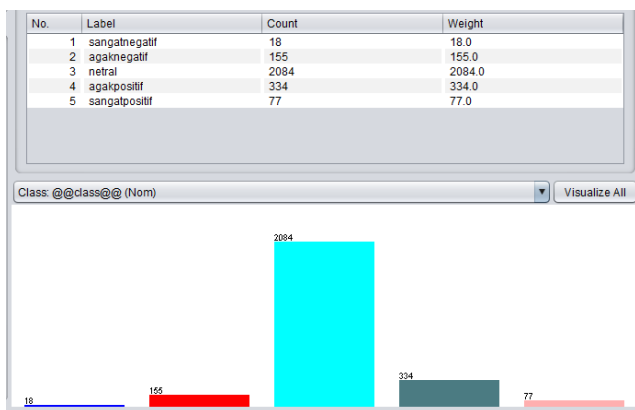
dimaksudkan agar tingkat kepercayaan dalam menilai sebuah kalimat itu positif maupun negative sangat tinggi dan dapat dipercaya.

Dari hasil preprocessing ini, terdapat data dengan nilai sangat negatif sebanyak 18 data, yang agak negatif 55 data, yang netral ada 2084 data, yang agak positif 334 data dan sangat positif ada 77 data (Gambar 2). Jika dipresentasikan dan mengabaikan data netral yang terpaut sangat jauh perbedaannya, maka akan terlihat grafik pada (Gambar 1).



Gambar 3. Persentase Pandangan Masyarakat

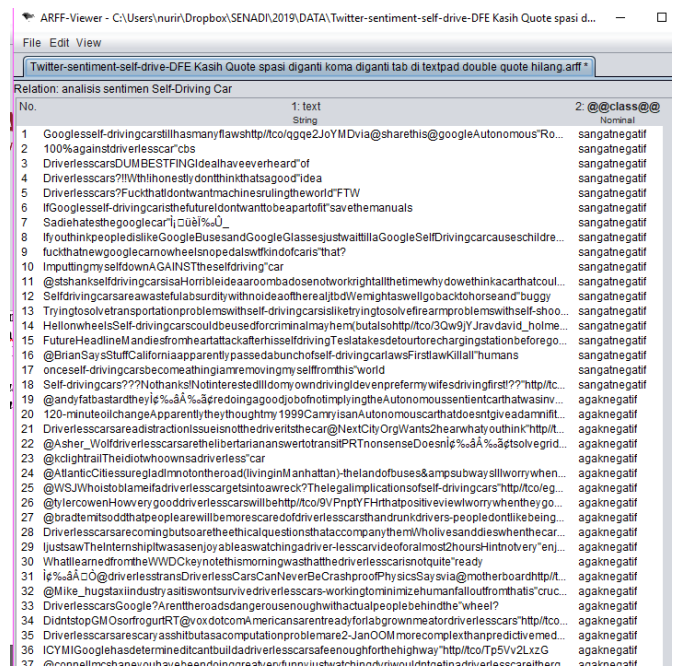
Selanjutnya pada field Golden dikarenakan semua nilainya sama yaitu sudah final, maka field Golden akan kita hapus saja. Kemudian yang akan kita hapus selanjutnya adalah sentimen confidence karena di sini kita menggunakan sentimen confidence dengan nilai 1 atau nilai yang terbaik, otomatis isinya hanya ada satu. selanjutnya pada data masih berupa angka atau atribut masih berupa angka, untuk itu perlu diubah dari angka satu menjadi sangat negative, angka 2 menjadi agak negative, angka 3 menjadi netral, angka 4 menjadi agak positif dan angka 5 menjadi sangat positif.



Gambar 4. Data setelah dilakukan preprocessing.

C. Mengubah data ke dalam format .arff

Setelah data dilakukan preprocessing, selanjutnya data diubah ke dalam format .arff. Proses ini dapat dilakukan di weka dengan masuk ke menu tools kemudian arff viewer (Gambar 3).



Gambar 5. Tampilan Arff Viewer pada WEKA

Bentuk data yang sudah terkonversi menjadi format .arff harus relasi, atribut dan data. Relasi diisi judul datanya, atribut diisi tipe data dan class atributnya, kemudian data harus diisi datanya. Data harus dipisahkan menggunakan tab di tiap kata dan harus berada diantara single quote ('). File yang sudah diubah menjadi .arff dapat dilihat sebagai berikut :

```
@attribute text string
@attribute @@class@@
{sangatnegatif,agaknegatif,netral,agakpositif,sangatpositif}
@data
'Audi is the first carmaker to
get a license from
Nevada DMV to test
automated vehicles audi ces
"cartech",netral"
'At the SAE On-Road Automated
Vehicle Standards Committee
Open "Meeting",netral"
'How is an automated vehicle
supposed to behave in
an emergency?
"AutoAuto",netral"
```

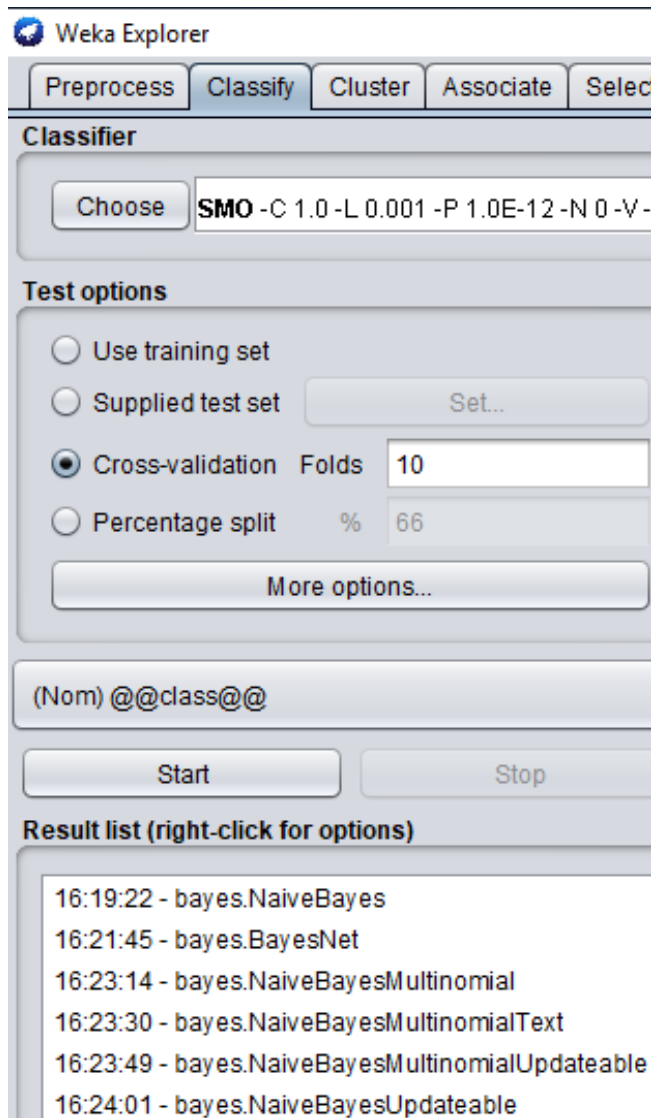
D. Filterisasi menggunakan StringToWordVector

Data yang sudah berformat .arff siap untuk diproses menggunakan tools WEKA. Filter yang dipakai adalah unsupervised learning, yaitu StringToWordVector. Kemudian selanjutnya akan terlihat visual seperti pada (Gambar 1)

E. Klasifikasi

Klasifikasi dilakukan pada tab Classify dan memilih metode yang ingin dipakai. Pada tab Classify, akan dipilih metode metode klasifikasi Naïve Bayes, Bayes Network, Naïve Bayes Multinomial, Naïve Bayes Multinomial Text, Naïve Bayes Multinomial Updateable, dan Naïve Bayes Updateable (Gambar 4).

Pengujian yang dilakukan adalah 10 Fold Cross Validation, yang pengetesan dengan membagi data menjadi 10% data testing dan 90% data training. Pada pengujian ini, data akan secara bergantian sebagai data training dan data testing hingga terbagi rata dan adil.

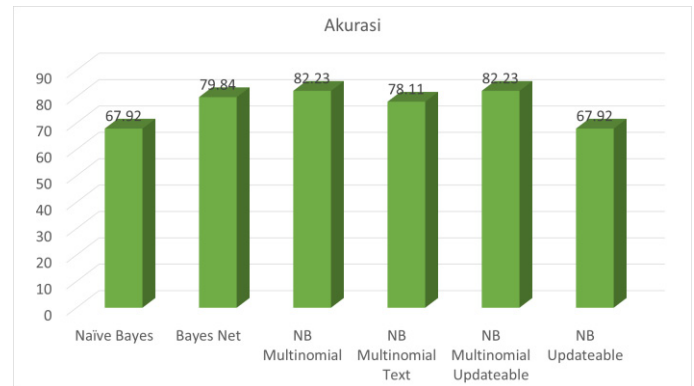


Gambar 6. Metode Klasifikasi

IV. HASIL DAN PEMBAHASAN

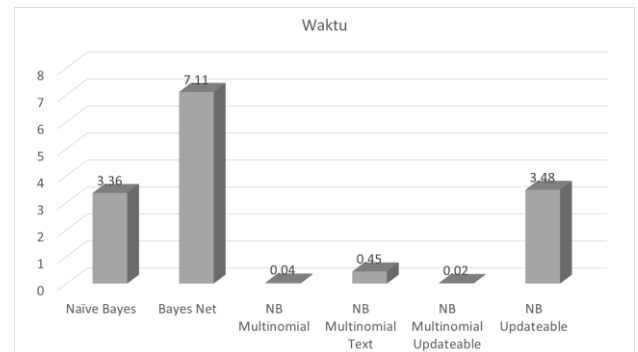
Berdasarkan klasifikasi menggunakan masing-masing metode, metode Naïve Bayes Multinomial menghasilkan akurasi yang paling tinggi yaitu sebesar 82,23%. Kemudian dilanjutkan dengan Bayes Network dengan akurasi sebesar 79,84 %, akurasi selanjutnya dengan menggunakan metode Naïve Bayes Multinomial Text sebesar 78,11%, dan yang terakhir adalah Naïve Bayes 67,92% (Gambar 5).

Metode pembaharuan Naïve Bayes dan Naïve Bayes Multinomial (Naïve Bayes Updateable dan Naïve Bayes Multinomial Updateable) ternyata tidak memberikan banyak perubahan dengan memberikan akurasi yang sama sebelum dilakukan perubahan (Gambar 5).



Gambar 7. Hasil Akurasi (%)

Waktu yang dibutuhkan masing-masing metode juga berbeda, waktu yang terbaik dalam klasifikasi adalah Metode Naïve Bayes Multinomial Updateable 0,02 detik, kemudian dilanjutkan dengan Naïve Bayes Multinomial 0,04 detik, Naïve Bayes Multinomial Text 0,45 detik, Naïve Bayes 3,36 detik, dan Bayes Network 7,11 detik.



V. PENUTUP

Kesimpulan yang dapat diambil dalam penelitian ini adalah dalam melakukan klasifikasi Analisis Sentimen khususnya data berupa text, Naïve Bayes Multinomial, Bayes Net dan Naïve Bayes Multinomial Text lebih baik dibandingkan Naïve Bayes umum. Selain akurasinya lebih tinggi, waktu yang dibutuhkan juga lebih optimal, hanya untuk Bayes Net waktu yang dibutuhkan jauh lebih lama dibandingkan yang lainnya.

UCAPAN TERIMA KASIH

Terima kasih kepada seluruh civitas akademika Universitas PGRI Yogyakarta, terima kasih kepada anak dan istri yang selalu memberikan semangat, serta terima kasih kepada seluruh peneliti yang menjadi inspirasi dalam artikel ini.

DAFTAR PUSTAKA

- [1] M. Bojarski *et al.*, "End to End Learning for *Self-driving cars*," *arXiv:1604.07316 [cs]*, Apr. 2016.
- [2] N. J. Goodall, "Can you program ethics into a *self-driving car*?," *IEEE Spectrum*, vol. 53, no. 6, pp. 28–58, Jun. 2016.
- [3] "'Self-Driving-Car', Masa Depan Big Data di Depan Mata," *idBigData*, 10-Jul-2017. .
- [4] Google, *Self-driving car Test: Steve Mahan*. .
- [5] "BMW: Bersama Apple Bikin Mobil Self Driving," *Car Review Indonesia*. [Online]. Available: <https://carreview.id/news/bmw-bersama-apple-bikin-mobil-self-driving/8916>. [Accessed: 23-Apr-2019].
- [6] M. König and L. Neumayr, "Users' resistance towards radical innovations: The case of the *self-driving car*," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 44, pp. 42–52, Jan. 2017.
- [7] A. H. Mirza, "APPLICATION OF NAIVE BAYES CLASSIFIER ALGORITHM IN DETERMINING NEW STUDENT ADMISSION PROMOTION STRATEGIES," *I*, vol. 1, no. 1, pp. 14–28, Mar. 2019.
- [8] M. Idris, "IMPLEMENTASI DATA MINING DENGAN ALGORITMA NAÏVE BAYES UNTUK MEMPREDIKSI ANGKA KELAHIRAN," *Pelita Informatika: Informasi dan Informatika*, vol. 18, no. 1, pp. 160–167, Apr. 2019.
- [9] G. Isabelle, W. Maharani, and I. Asror, "Analysis on Opinion Mining Using Combining Lexicon-Based Method and Multinomial Naïve Bayes," presented at the 2018 International Conference on Industrial Enterprise and System Engineering (ICoIESE 2018), 2019.
- [10] N. Saputra, "ANALISIS SENTIMEN DENGAN PREPROCESSING KATA (SENTIMENT ANALISYS WITH LEXICON PREPROCESSING)," *I*, vol. 7, no. 1, pp. 45–57, 2018.
- [11] N. Saputra, "Analisis Sentimen Mahasiswa Terhadap Universitas."
- [12] L. Jiang, L. Zhang, L. Yu, and D. Wang, "Class-specific attribute weighted naive Bayes," *Pattern Recognition*, vol. 88, pp. 321–330, Apr. 2019.
- [13] W. Xu, L. Jiang, and L. Yu, "An attribute value frequency-based instance weighting filter for naive Bayes," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 31, no. 2, pp. 225–236, Mar. 2019.
- [14] L. Jiang, L. Zhang, C. Li, and J. Wu, "A Correlation-Based Feature Weighting Filter for Naive Bayes," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 2, pp. 201–213, Feb. 2019.
- [15] G. H. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers," *arXiv:1302.4964 [cs, stat]*, Feb. 2013.
- [16] D. Soni, "Introduction to Bayesian Networks," *Towards Data Science*, 08-Jun-2018. [Online]. Available: <https://towardsdatascience.com/introduction-to-bayesian-networks-81031eed94e>. [Accessed: 11-May-2019].
- [17] M. S. Geraldi and E. Ghisi, "Short-term instead of long-term rainfall time series in rainwater harvesting simulation in houses: An assessment using Bayesian Network," *Resources, Conservation and Recycling*, vol. 144, pp. 1–12, May 2019.
- [18] S. Kadam, A. Gala, P. Gehlot, A. Kurup, and K. Ghag, "Word Embedding Based Multinomial Naive Bayes Algorithm for Spam Filtering," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018, pp. 1–5.
- [19] "Sentiment *Self-driving cars* - dataset by crowdflower," *data.world*. [Online]. Available: <https://data.world/crowdflower/sentiment-self-driving-cars>. [Accessed: 10-May-2019].