

Implementasi Metode Klasifikasi Naive Bayes Pada Sistem Analisis Opini Pengguna Twitter Berbasis Web

Apif Supriadi¹, Fatmasari²

Abstract— *Development of social media which is the result of technological development is an inseparable part of people's lives. Social media is a place where ordinary people express their feelings and opinions about something that concerns them. In knowing the direction of public sentiment, surveys are usually done online or offline, this sentiment analysis system will facilitate and speed up the process of knowing the direction of public sentiment, in the case of research. This uses data from Twitter social media called tweets or tweets, web-based sentiment analysis system that will classify tweets into 3 (three) types of sentiments, namely positive, neutral and negative, then make a percentage to make it easier to see the direction of public sentiment. In classifying this system uses the Naive Bayes Classifier method and displays it in a web interface with the PHP programming language and uses the Application Programming Interface (API) to get data from Twitter.*

Intisari — Saat ini perkembangan media sosial yang merupakan hasil dari perkembangan teknologi menjadi bagian tak terpisahkan dari kehidupan masyarakat. Media sosial menjadi tempat masyarakat biasa mengutarakan berbagai perasaan dan opininya tentang suatu hal yang jadi perhatian mereka, dalam mengetahui arah sentimen masyarakat biasanya dilakukan survei baik secara online atau offline, sistem analisis sentimen ini akan memudahkan dan mempercepat proses mengetahui arah sentimen publik, dalam kasus penelitian ini menggunakan data dari media sosial Twitter yang disebut dengan tweets atau cuitan, sistem analisis sentimen berbasis web yang akan mengklasifikasikan cuitan kedalam 3 (tiga) jenis sentimen yaitu positif, netral dan negatif lalu melakukan persentasenya agar mempermudah melihat arah sentimen publik. Dalam melakukan klasifikasinya sistem ini menggunakan metode Naive Bayes Classifier dan menampilkannya dalam antarmuka web dengan bahasa pemrograman PHP dan menggunakan Application Programming Interface (API) dalam mendapatkan data dari Twitter.

Kata Kunci — *Application Programming Interface (API), Sistem analisis sentimen, Twitter, Naive Bayes Classifier, Web*

I. PENDAHULUAN

Internet telah menjadi bagian penting dalam kehidupan kita sehari-hari hampir semua pekerjaan kita hari ini tidak pernah lepas dari koneksi jaringan internet artinya dalam

setiap aktivitas, Kita pasti selalu terhubung dengan internet hal ini pula banyak memunculkan kehidupan sosial baru di masyarakat, tumbuhnya internet juga berbanding lurus dengan sistem demokrasi yang menuntut kebebasan dalam berpendapat di belahan dunia manapun, Internet pun tidak mengenal batas - batas privasi dan negara.

Penggunaan internet banyak dipakai oleh masyarakat khususnya di Indonesia untuk mengakses media sosial dan melakukan komunikasi pesan singkat, dalam penggunaannya media sosial menjadi alat untuk mencurahkan atau mengekspresikan opini dan cerita pengalaman pengguna, hal ini bisa diistilahkan sebagai opini.

Karena banyaknya masyarakat Indonesia yang menggunakan media sosial, opini masyarakat Indonesia banyak bertebaran di media sosial, opini tersebut pun beragam dari mulai pengalaman dalam memakai suatu produk atau berkomentar soal pertandingan olahraga atau memberikan opininya terkait isu terkini di Indonesia sampai ke mengekspresikan pandangan politik di Indonesia.

Media sosial yang sering diakses oleh masyarakat Indonesia dan dunia itu diantaranya adalah Facebook, Instagram dan Twitter, dalam penelitian ini penulis akan menitikberatkan pada media sosial Twitter. Jumlah pengguna Twitter di Indonesia merupakan pangsa pasar yang menjanjikan. Maka berbagai produsen mulai dari kelas kecil hingga besar berlomba-lomba mengelola potensi ekonomi besar ini agar produk - produk mereka laku di pasaran sekurang-kurangnya menjadi referensi. Bukan hanya kalangan bisnis saja yang berlomba mengelola opini publik lewat media sosial ini tokoh publik dan partai politik juga memanfaatkan potensi ini sebagai acuan mereka dalam hal mengelola citra di masyarakat. Analisis opini yang merupakan bagian dari opinion mining. Analisis opini dilakukan untuk melihat pendapat terhadap sebuah masalah atau dapat juga digunakan untuk identifikasi kecenderungan hal di pasar.

Saat ini Twitter merupakan sebuah indikator yang baik untuk memberikan pengaruh dalam penelitian. Namun masih belum banyak aplikasi dan metode analisa sentimen yang dikembangkan untuk bahasa Indonesia. Faktor-faktor keuntungan tersebut mendorong perlunya dilakukan penelitian analisis opini terhadap dokumen berbahasa Indonesia. Analisis sentimen dalam penelitian ini merupakan proses klasifikasi dokumen tekstual ke dalam 3 kelas, yaitu kelas sentimen positif, netral dan negatif. Data opini diperoleh dari media sosial Twitter berdasarkan query dalam Bahasa Indonesia. Penelitian ini bertujuan untuk menentukan sentimen publik terhadap objek tertentu yang disampaikan di Twitter dalam Bahasa Indonesia, sehingga

¹Program Studi Teknik Informatika STMIK STMIK Antar Bangsa, Jl. HOS Cokroaminoto Kawasan Bisnis CBD Ciledug, Blok A5, No.29-36, Karang Tengah, Kota Tangerang; Telp. (021) 5098-6099; e-mail: afs.supriadi@gmail.com

²Program Studi Sistem Informasi STMIK STMIK Antar Bangsa, Jl. HOS Cokroaminoto Kawasan Bisnis CBD Ciledug, Blok A5, No.29-36, Karang Tengah, Kota Tangerang; Telp. (021) 5098-6099; e-mail: fsarie@gmail.com

membantu usaha untuk melakukan riset pasar atas opini publik.

Dari beberapa uraian yang dikemukakan pada latar belakang, maka dapat diidentifikasi masalah - masalah sebagai berikut : 1) Kurangnya pemanfaatan opini publik di media sosial untuk mengetahui kecenderungan hal yang ada di pasar, 2) analisis opini banyak dibutuhkan oleh tokoh publik, perusahaan, pemerintahan, dan lain-lain, 3) Salah satu kendala utama dalam analisis sentimen adalah terbatasnya sistem yang bisa menganalisis sentimen berbahasa Indonesia secara khusus.

Tujuan dari penelitian ini adalah membuat sistem analisis sentimen yang bisa menjadi acuan dalam penentuan kebijakan baik pemerintah atau umum dan untuk mengetahui kapan dan apa penyebab perubahan sentimen di media sosial Twitter.

Dalam penelitian ini, penulis sedikit banyak terinspirasi dan mereferensi dari penelitian – penelitian sebelumnya yang berkaitan dengan latar belakang masalah pada penelitian ini, antara lain yaitu penelitian yang dilakukan oleh Ismail Sunni dan Dwi Hendratmo Widyantoro dengan judul Analisis sentimen dan ekstraksi topik penentu sentimen pada opini terhadap tokoh publik. Solusi yang diberikan dalam penelitian tersebut yaitu penggunaan sistem F3 yang dibangun dengan memiliki empat buah subsistem yaitu subsistem ekstraksi data mentah, subsistem model analisis sentimen, subsistem klasifikasi sentimen, dan subsistem ekstraksi topik. Masukan dari luar sistem adalah nama tokoh dan data dari Twitter. Nama tokoh dijadikan sebagai filter untuk mendapatkan tweet yang mengandung nama tokoh yang bersangkutan. Sedangkan kesimpulan yang bisa diambil dari penelitian tersebut adalah metode praproses yang dicoba dalam penelitian ternyata tidak memiliki pengaruh yang signifikan terhadap akurasi sistem. Penyebab utamanya adalah domain tweet tokoh masyarakat yang sering berisi judul berita. Sedangkan penggunaan faktor peluruhan dalam pencarian topik ternyata mampu menaikkan jumlah topik yang berhasil diekstrak. Namun masih tidak terlalu baik ketika berhadapan dengan topik yang muncul di hampir semua waktu [1].

Pada penelitian ini penulis juga mengacu pada penelitian yang dilakukan oleh Noviah Dwi Putranti dan Edi Winarko dengan judul penelitian Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. Penelitian analisis sentimen ini dilakukan untuk mengetahui sentimen publik mengenai sesuatu dengan menggunakan pendekatan dalam machine learning yang dikenal dengan nama Support Vector Machine dan Maximum Entropy Part of Speech Tagging yang dikhususkan pada dokumen teks berbahasa Indonesia dengan fitur n gram [2].

Penulis juga mengacu pada Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen di Twitter yang dilakukan oleh Nurun Mochammad Shiddieqy Hadna, dkk. Penelitian ini dilakukan untuk mengetahui bagaimana perbandingan dua metode yang bisa digunakan untuk melakukan proses analisis sentimen disertai tahapan yang akan dilalui. Metode yang dibandingkan adalah Naive Bayes Classifier dan Support Vector Machine. Dari beberapa penelitian sebelumnya, bisa dilihat bahwa hasil terbaik dapat

diraih menggunakan metode Support Vector Machine. Namun, untuk tetap mendapatkan hasil kualifikasi terbaik, maka lebih baik dilakukan pengujian yang bersifat komparatif dengan beberapa metode klasifikasi pada penelitian yang sedang dilakukan [3].

Penelitian selanjutnya yang menjadi acuan penulis adalah Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter yang dilakukan oleh Ghulam Asrofi Guntoro. 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. Dari penelitian yang telah dilakukan, maka dapat ditarik kesimpulan bahwa Analisis Sentimen dapat digunakan untuk mengetahui sentimen masyarakat khususnya netizen Twitter terhadap calon Gubernur DKI Jakarta 2017. Tujuannya membantu masyarakat menentukan sentimen yang terdapat pada tweet opini Bahasa Indonesia yang ada di Twitter [4].

Penulis juga mengacu pada penelitian yang dilakukan oleh Ahmad Fathan Hidayatullah dan Azhari SN dengan judul Analisis Sentimen Dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter. Penelitian ini telah berhasil membangun model untuk melakukan klasifikasi tweet berdasarkan sentimen dan kategori dengan Naive Bayes Classifier. Hasil pengujian pada aplikasi yang dibangun dan pada tools Rapidminer memperlihatkan bahwa akurasi dengan fitur term frequency memberikan hasil akurasi yang lebih baik daripada akurasi dengan fitur TF-IDF. Metode Support Vector Machine menghasilkan akurasi performansi yang lebih baik daripada metode Naive Bayes. Penggunaan metode Support Vector Machine dan Naive Bayes sama-sama memiliki hasil akurasi yang cukup baik untuk klasifikasi tweet [5].

II. TINJAUAN PUSTAKA

A. Sistem Analisis Sentimen

Untuk mengetahui pengertian dari sistem analisis sentimen bisa diuraikan seperti berikut :

1. Sistem

Suatu sistem adalah suatu jaringan kerja dari prosedur-prosedur yang saling berhubungan, berkumpul bersama-sama untuk melakukan suatu kegiatan atau untuk menyelesaikan suatu sasaran tertentu [6].

2. Analisis

- a) Menurut Kamus Besar Bahasa Indonesia (KBBI) kata analisis mempunyai arti, diantaranya sebagai berikut : Penyelidikan terhadap suatu peristiwa (Karangan, perbuatan dan sebagainya) untuk mengetahui keadaan sebenarnya (sebab-musabab, duduk perkaranya dan sebagainya).
- b) Penguraian suatu pokok atas berbagai bagiannya dan penelaahan bagian itu sendiri serta hubungan antar bagian untuk memperoleh pengertian yang tepat dan pemahaman arti keseluruhan.

3. Sentimen

Sentimen atau opini masyarakat semakin bertambah luas dan bebas diungkapkan di berbagai media. Sentimen dapat menjadi potensi besar bagi perusahaan yang ingin mengetahui umpan balik (feedback) dari masyarakat terhadap merek dagang mereka. Merek telah dianggap sebagai salah satu aset tidak berwujud (intangible asset) yang paling berharga dan manajemen merek adalah prioritas penting bagi pihak manajemen perusahaan atau organisasi [7].

Dapat disimpulkan dari uraian diatas bahwa Sentiment Analysis adalah merupakan perpaduan dari data mining dan text mining, atau sebuah cara yang digunakan untuk mengolah berbagai opini yang diberikan oleh konsumen atau para pakar melalui berbagai media, mengenai sebuah produk, jasa ataupun sebuah instansi. Sentiment analysis merupakan sebuah metode yang digunakan untuk memahami, mengekstrak data opini, dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah sentiment yang terkandung dalam sebuah opini. Pada Sentiment analysis terdiri dari 3 jenis opini, yaitu opini positif, opini negatif dan opini netral, sehingga dengan sentiment analysis perusahaan atau instansi yang terkait dapat mengetahui respon masyarakat terhadap suatu pelayanan atau produk, melalui feedback masyarakat maupun para ahli. Sentiment mengacu pada fokus topik tertentu, pernyataan pada suatu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subjek yang berbeda, oleh karena itu pada beberapa penelitian, terutama pada review produk, pekerjaan didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses Sentiment Analysis [8].

B. Tahapan analisis sentimen

Untuk mendapatkan hasil pengujian yang maksimal dalam menganalisa sentimen yang berupa text diperlukan beberapa tahapan yang perlu dilewati, Berikut beberapa tahapannya :

1. Pengumpulan dan Pelabelan data

Tahap pertama untuk melakukan proses analisis sentimen adalah pengumpulan data. Dalam penelitian ini, data diambil dari Twitter. Pengambilan data dari Twitter dengan REST API, Cukup mudah dilakukan karena Twitter sudah menyediakan API (Application Programming Interface) yang ditujukan kepada pengembang sistem untuk mempermudah pengambilan data dari Twitter.

Setelah data berhasil dikumpulkan menjadi sebuah dataset, tahap selanjutnya adalah pelabelan. Pelabelan di sini ditujukan untuk membagi data ke dalam beberapa kelas sentimen yang akan digunakan dalam penelitian. Jumlah kelas sentimen yang banyak digunakan adalah dua dan tiga kelas, yaitu negatif, netral dan positif. Tujuan dari proses pelabelan ini adalah membagi dataset menjadi 2 bagian, yaitu menjadi training data dan testing data. Training data adalah data yang digunakan untuk melatih sistem agar mampu mengenali pola yang sedang dicari, sedangkan testing data adalah data yang digunakan untuk

menguji hasil pelatihan yang sudah dilakukan. Untuk melakukan pembagian, salah satu metode yang bisa digunakan adalah n-fold cross validation [3].

2. Preprocessing

Setelah melewati pelabelan data, selanjutnya adalah tahapan preprocessing, tahapan dimana data disiapkan agar menjadi data yang siap untuk dianalisis. Ada beberapa tahap dalam preprocessing ini, antara lain tokenization, cleansing dan filtering. Berikut ini adalah penjelasan secara rinci tentang tahap preprocessing, yaitu :

a) Tokenization

Sebuah proses yang dilakukan untuk memotong atau memecah kalimat menjadi beberapa bagian atau kata Manning et al [3]. Hasil dari pemotongan ini disebut dengan token. Pada beberapa kasus, proses tokenisasi juga dilakukan dengan membuang tanda baca yang tidak diperlukan. Ada beberapa model tokenization yang bisa digunakan, yaitu unigram, bigram, trigram dan ngram [3] .

b) Cleansing

Suatu tahap di mana karakter maupun tanda baca yang tidak diperlukan dibuang dari teks. Contoh karakter yang dibuang adalah tanda seru, tanda tanya, koma dan titik [3].

c) Filtering

Merupakan tahap menghilangkan kata-kata yang muncul dalam jumlah besar, namun dianggap tidak memiliki makna (*stopwords*). Pada dasarnya, stopwords list adalah sekumpulan kata - kata yang banyak digunakan dalam berbagai bahasa. Alasan penghapusan stopwords dalam banyak program aplikasi yang berkaitan dengan text mining adalah karena penggunaannya yang terlalu umum, sehingga pengguna dapat berfokus pada kata-kata lain yang jauh lebih penting (Ganesan dalam [3]).

3. Pembobotan kata

TF-IDF (Term Frequency-Inverse Document Frequency). Term Frequency-Inverse Document Frequency adalah sebuah metode pembobotan yang menggabungkan dua konsep, yaitu Term Frequency dan Document Frequency (Asrofi, 2015). Term Frequency adalah konsep pembobotan dengan mencari seberapa sering (frekuensi) munculnya sebuah term dalam satu dokumen (Huang & Wu dalam [3]).

Sedangkan Document Frequency adalah banyaknya jumlah dokumen di mana sebuah term itu muncul (Huang & Wu dalam [3]). Semakin kecil frekuensi kemunculannya, maka semakin kecil pula nilai bobotnya. Ketika proses perhitungan term frequency, semua kata di dalamnya dianggap sama pentingnya. Namun, terdapat kata yang sebenarnya kurang penting dan tidak perlu diperhitungkan seperti “di-”, “ke-”, “dan” dan lain sebagainya. Oleh sebab itu, kata-kata yang kurang penting tersebut perlu dikurangi bobotnya dan menambah bobot kata penting lainnya. Inilah ide dasar mengapa diperlukan stopwords.

C. Metode Klasifikasi Naive Bayes

Supriyanto [10] menjelaskan bahwa Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Bayesian classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Bayesian classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar, Metode Bayes merupakan pendekatan statistik untuk melakukan inferensi induksi pada persoalan klasifikasi. Pertama kali dibahas terlebih dahulu tentang konsep dasar dan definisi pada Teorema Bayes, kemudian menggunakan teorema ini untuk melakukan klasifikasi dalam Data Mining.

D. Algoritma Naive Bayes

Algoritma Naive Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan Naive dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi Naive Bayes diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Menurut Saleh [11] Persamaan dari teorema Bayes adalah :

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)} \quad \dots\dots\dots (1)$$

Keterangan :

X = Data dengan class yang belum diketahui
H = Hipotesis data X merupakan suatu class spesifik
P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (porteriori prob.)
P(H) = Probabilitas hipotesis H (prior prob.)
P(X|H) = Probabilitas X berdasarkan kondisi tersebut
P(X) = Probabilitas dari X

Untuk menjelaskan teorema Naive Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, Menurut Saleh [11] teorema bayes di atas disesuaikan sebagai berikut :

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(C|F_1 \dots F_n)} \quad \dots\dots\dots (2)$$

Dimana Variabel C merepresentasikan kelas, sementara variabel $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C

(disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik karakteristik sampel secara global (disebut juga evidence). Karena itu, Menurut Saleh [11] rumus diatas dapat pula ditulis secara sederhana sebagai berikut :

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{evidence}} \quad \dots\dots\dots (3)$$

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $(C|F_1, \dots, F_n)$ Menurut Saleh [11] menggunakan aturan perkalian sebagai berikut :

$$\begin{aligned} P(C|F_1 \dots F_n) &= P(C)P(F_1, \dots, F_n|C) \\ &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing masing petunjuk ($F_1, F_2, \dots, F_n$) saling bebas (independen) satu sama lain. Dengan asumsi tersebut, Menurut Saleh [11] maka berlaku suatu kesamaan sebagai berikut:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i)$$

Untuk $i \neq j$ Sehingga

$$P(F_i|C, F_j) = P(F_i|C)$$

Dari persamaan diatas dapat disimpulkan bahwa asumsi independensi naif tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran $P(C|F_1, \dots, F_n)$ Menurut Saleh [11] dapat disederhanakan menjadi :

$$\begin{aligned} P(C|F_1 \dots F_n) &= P(C)P(F_1|C)P(F_2|C)P(F_3|C) \dots \\ &= P(C) \prod_{i=1}^n P(F_i|C) \end{aligned}$$

Adapun alur dari metode Naive Bayes adalah sebagai berikut :

1. Baca data training
2. Hitung Jumlah dan probabilitas, namun apabila data numerik maka:
 - a. Cari nilai mean dan standar deviasi dari masing masing parameter yang merupakan data numerik.
 - b. Cari nilai probabilitistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
1. Mendapatkan nilai dalam tabel mean, standart deviasi dan probabilitas.

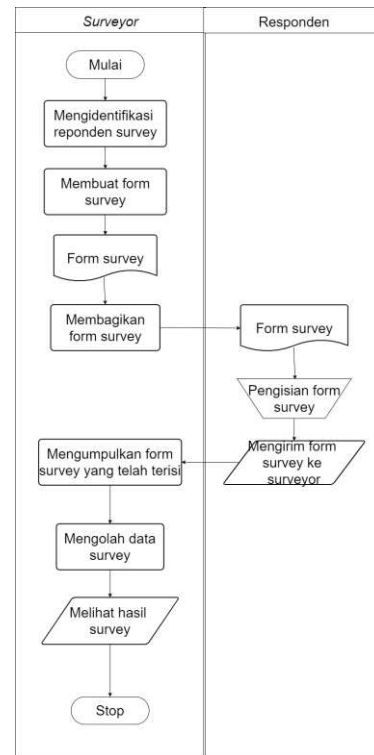
E. Twitter

Twitter merupakan jenis situs jejaring sosial pertemanan yang memungkinkan para penggunanya mendapatkan relasi dengan mendaftarkan dirinya di situs tersebut. Kehadiran Twitter Semakin meramaikan persaingan situs-situs social networking yang sudah ada sebelumnya, seperti Friendster, Facebook, Plurk dan lain sebagainya. Twitter Didirikan oleh tiga orang, yaitu Jack Dorsey, Biz Stone, dan Evan Williams pada bulan Maret 2006, dan baru diluncurkan pada Juli di tahun yang sama. Twitter Merupakan jejaring sosial dan microblogging yang memfasilitasi pengguna, dapat memberikan update (pembaharuan) informasi tentang diri, hingga kegiatan bisnis. Setiap menulis status pada Twitter, status tersebut disebut sebagai Tweets. Apabila jumlah Tweets Berjumlah 50, maka itu artinya pengguna telah menulis status sebanyak 50 kali. Tweets Merupakan penulisan teks berbasis 140 karakter. Jadi, jumlah maksimal karakter yang dituliskan sebagai status hanya terbatas maksimal 140 karakter. Tweets Dapat ditampilkan pada profil, atau dapat mengomentari status dari teman. Keistimewaan Tweets Adalah dapat mengirimkan melalui Twitter Melalui internet, SMS (*Short Message Service*), atau aplikasi yang lain. Tentunya, pengiriman Tweet Melalui SMS akan berpengaruh pada penggunaan ponsel [9].

III. METODE PENELITIAN

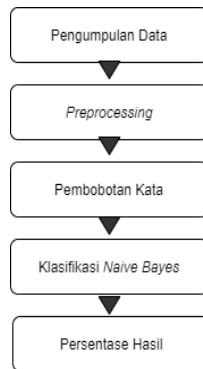
Analisis sistem yang sedang berjalan didefinisikan sebagai penguraian dari suatu sistem untuk mengidentifikasi dan mengevaluasi permasalahan yang terjadi, adapun dalam mengetahui opini atau sentimen publik terhadap suatu topik atau sebuah produk masih banyak mengandalkan survei baik offline atau online, dan kotak kritik saran. Tentunya cara tersebut masih layak untuk dipakai untuk bisa mengetahui opini atau sentimen publik terhadap suatu topik atau produk, tetapi cara tersebut mengandalkan kesukarelaan dari publik untuk mengisi atau menulis kritik saran dan opininya yang tentunya tidak semua orang berkeinginan membantu, ini menjadi persoalan sendiri ditambah dengan memerlukan waktu sedikit lebih lama dan mengeluarkan biaya yang tidak sedikit juga terkendala masalah dimana sukarelawan terkadang tidak berintegritas dalam mengisi atau menuliskan masukannya menjadikan cara tersebut masih kurang efektif dalam mengetahui opini dan sentimen publik.

Sekarang, media sosial menjadi platform dimana semua orang bisa mengutarakan opini atau sentimennya terhadap suatu topik yang diperhatikan oleh mereka hal ini bisa menjadi andalan dalam mengetahui arah opini atau sentimen publik.



Gbr 4. Flowchart survey online atau offline untuk mengetahui opini publik

Untuk mendapatkan hasil analisis baik melakukan klasifikasikan kedalam 3 jenis sentimen untuk setiap cuitan dan menghasilkan persentase hasil dari masing - masing jenis sentimen, sistem tersebut telah melalui beberapa tahapan, alur kerja dari sistem analisis sentimen diawali dengan pengumpulan data yang didapatkan setelah menerima response dari aplikasi Twitter lalu disambung ke proses preprocessing yaitu tahapan dimana sistem melakukan tokenisasi, cleansing dan filtering pada text yang merupakan cuitan dari pengguna twitter, selanjutnya ialah pembobotan kata pada setiap cuitan yang diolah dan pada proses terakhir dilakukan klasifikasi naive bayes untuk menghitung nilai probabilitas di tiap text dan memasukkannya dalam 3 (tiga) jenis sentimen atau setiap text yang dihitung probabilitasnya akan memiliki 3 (tiga) peluang setelah proses analisis tadi selesai maka selanjutnya sistem akan menghitung persentase untuk masing - masing sentimen dimana jumlah data tiap sentimen dibagi total jumlah data keseluruhan dikalikan 100 maka hasilnya adalah persentase dari setiap masing - masing sentimen. Tahap-tahap yang akan dilakukan dalam penelitian ini diilustrasikan pada gambar 3.



Gbr 3. Tahapan untuk hasil analisis sentiment

IV. PEMBAHASAN DAN HASIL

A. Sistem Usulan

Dalam penelitian ini ingin mengusulkan agar dalam proses mengetahui arah sentimen publik terhadap suatu topik tertentu menggunakan sistem analisis sentimen dengan menggunakan data cuitan atau tweet dari media sosial twitter. media sosial menjadi platform dimana semua orang bisa mengutarakan opini atau sentimennya terhadap suatu topik yang diperhatikan oleh mereka hal ini bisa menjadi andalan dalam mengetahui arah opini atau sentimen publik.



Gbr 5. Sistem Analisis Opini Publik

B. Implementasi

Implementasi adalah penerapan cara kerja sistem berdasarkan hasil analisa dan juga perancangan yang telah dibuat sebelumnya ke dalam suatu bahasa pemrograman tertentu. Sistem ini berbasis web menggunakan bahasa pemrograman PHP sebagai backend, HTML dan CSS sebagai frontend dan MySql sebagai database.

1. Implementasi Database

Pembuatan database dilakukan dengan menggunakan aplikasi DBMS MySQL dengan nama database sentimentdb.

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	id	int(11)			No	None		AUTO_INCREMENT	Change Drop More
2	katakunci	varchar(100)	latin1_swedish_ci		No	None			Change Drop More
3	jumlahtweet	int(100)			No	None			Change Drop More
4	tipehasil	varchar(30)	latin1_swedish_ci		No	None			Change Drop More
5	tanggal	date			No	None			Change Drop More
6	positif	varchar(10)	latin1_swedish_ci		No	None			Change Drop More
7	negatif	varchar(10)	latin1_swedish_ci		No	None			Change Drop More
8	total	int(10)			No	None			Change Drop More

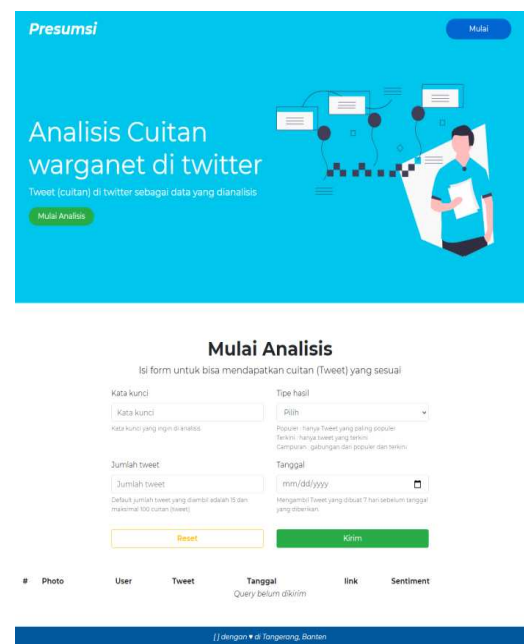
Gbr 6. Tabel database sentimentdb

2. Implementasi Antarmuka

Berikut ini adalah implementasi antarmuka aplikasi web analisis sentimen yang telah dibuat.

a) Halaman awal

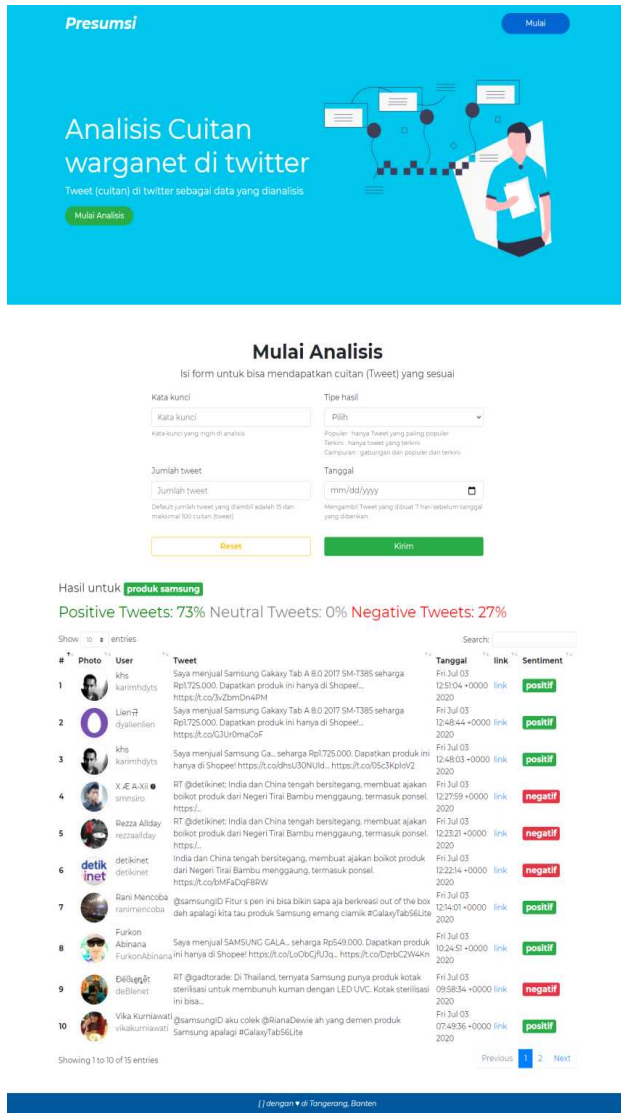
Halaman awal adalah halaman yang akan tampil pertama kali saat pengguna mengakses web analisis sentimen.



Gbr 7. Antarmuka Halaman awal

b) Halaman Hasil Analisis

Halaman hasil adalah halaman yang akan menampilkan hasil analisis sentimen berupa tabel cuitan dan hasil persentase dari 3 (tiga) jenis sentimen untuk memudahkan pengguna untuk mengetahui hasil analisis sentimen terhadap kata kunci yang telah diberikan sebelumnya.



Gbr 8. Antarmuka Halaman Analisis

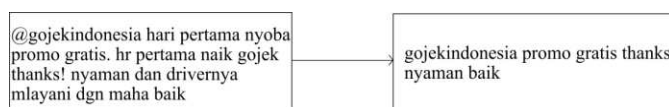
C. Analisis Opini Dengan Metode Klasifikasi Naive Bayes

Contoh kasus dari analisis opini dengan menggunakan metode klasifikasi naive bayes adalah sebagai berikut :

1. Tahap awal ialah pre-processing untuk mendapatkan data bersih. Tahapannya adalah case folding, tokenizing, stemming dan filtering. Berikut ini adalah data latih sebelum dan sesudah melewati proses pre-processing.



Gbr 9. Data Latih 1 Sebelum Dan Sesudah Pre-Processing



Gbr 10. Data Latih 2 Sebelum Dan Sesudah Pre-Processing

2. Dilakukan proses pembentukan fitur, perhitungan nilai probabilitas, dan menentukan probabilitas kelas. Berikut ini adalah hasil pembentukan fitur pada masing - masing data latih.

TABEL I
PEMBENTUKAN FITUR DATA LATIH

Data	Fitur	Kemunculan	Kelas Sentimen
D1	gojekindonesia	1	netral
	driver	1	netral
	gojek	1	netral
	jarang	1	negatif
	orderan	1	netral
	rugi	1	negatif
D2	gojekindonesia	1	netral
	promo	1	positif
	gratis	1	positif
	thanks	1	positif
	nyaman	1	positif
	baik	1	positif

3. Menghitung probabilitas kemunculan kata pada setiap kelas. Hasilnya adalah sebagai berikut :

TABEL II
FREKUENSI KEMUNCULAN KATA

Kelas Sentimen	Data (j)		Fd (C _j)	P(C _j)
	D1	D2		
Positif	0	1	1	1/2
Negatif	1	0	2	1/2
Netral	1	1	2	2/2

4. Setelah didapat probabilitas dari setiap kelas, selanjutnya menghitung probabilitas setiap fitur pada setiap kelas dengan hasil sebagai berikut :

TABEL III
PROBABILITAS TIAP FITUR PADA KELAS SENTIMEN

Data f(c _j)	Kelas Sentimen		
	Positif	Negatif	Netral
gojekindonesia	$\frac{0+1}{5+12} = \frac{1}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{1+1}{5+12} = \frac{2}{17}$
driver	$\frac{0+1}{5+12} = \frac{1}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{1+1}{5+12} = \frac{2}{17}$
gojek	$\frac{0+1}{5+12} = \frac{1}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{1+1}{5+12} = \frac{2}{17}$
jarang	$\frac{0+1}{5+12} = \frac{1}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$
pesan	$\frac{0+1}{5+12} = \frac{1}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{1+1}{5+12} = \frac{2}{17}$
rugi	$\frac{0+1}{5+12} = \frac{1}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$
promo	$\frac{1+1}{5+12} = \frac{2}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$
gratis	$\frac{1+1}{5+12} = \frac{2}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$
thanks	$\frac{1+1}{5+12} = \frac{2}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$
nyaman	$\frac{1+1}{5+12} = \frac{2}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$
baik	$\frac{1+1}{5+12} = \frac{2}{17}$	$\frac{0+1}{2+12} = \frac{1}{14}$	$\frac{0+1}{5+12} = \frac{1}{17}$

5. Menghitung V_{map} untuk menentukan probabilitas tertinggi dari masing-masing kelas berdasarkan dari proses pelatihan. Nilai probabilitas tertinggi merupakan sentimen dari data *tweet* atau cuitan tersebut.

a. V_{map} untuk sentimen positif

$$\begin{aligned} V_{map}(\text{positif}) &= P(\text{"positif"}) P(\text{"thanks"}|\text{positif}) P(\text{"gojek"}|\text{positif}) \\ P(\text{"baik"}|\text{positif}) \\ &= \frac{1}{2} \times \frac{2}{17} \times \frac{1}{17} \times \frac{2}{17} = \mathbf{0.000396981} \end{aligned}$$

b. V_{map} untuk sentimen negatif

$$\begin{aligned} V_{map}(\text{negatif}) &= P(\text{"negatif"}) P(\text{"thanks"}|\text{negatif}) P(\text{"gojek"}|\text{negatif}) \\ P(\text{"baik"}|\text{negatif}) \\ &= \frac{1}{2} \times \frac{1}{14} \times \frac{1}{14} \times \frac{1}{14} = \mathbf{0.0001789555} \end{aligned}$$

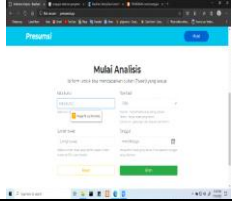
c. V_{map} untuk sentimen netral

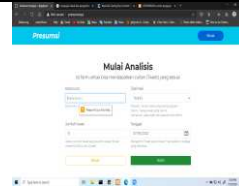
$$\begin{aligned} V_{map}(\text{netral}) &= P(\text{"netral"}) P(\text{"thanks"}|\text{netral}) P(\text{"gojek"}|\text{netral}) \\ P(\text{"baik"}|\text{netral}) \\ &= \frac{2}{2} \times \frac{1}{17} \times \frac{2}{17} \times \frac{1}{17} = \mathbf{0.000393588} \end{aligned}$$

D. Pengujian

Pada tahap ini akan dilakukan pengujian sistem yang bertujuan untuk menemukan kesalahan kesalahan atau kekurangan kekurangan pada perangkat lunak yang diuji. Pengujian bermaksud untuk mengetahui perangkat lunak yang dibuat sudah memenuhi kriteria yang sesuai dengan tujuan perancangan perangkat lunak tersebut. Pengujian aplikasi ini menggunakan metode pengujian *black box*. Pengujian *black box* berfokus pada persyaratan fungsional perangkat lunak.

TABEL IV
PENGUJIAN BLACKBOX PADA APLIKASI SISTEM ANALISIS SENTIMEN

No.	Skenario Pengujian	Hasil yang diharapkan	Kesimpulan
1	Mengosongkan semua form input langsung klik tombol "kirim" <i>Test Case :</i> 	Sistem akan menolak dan menampilkan "please fill out this field" pada input bagian kata kunci <i>Hasil yang diharapkan :</i> 	Valid
2	Mengisi form selain kata kunci lalu langsung klik tombol "kirim" <i>Test Case :</i>	Sistem akan menolak dan menampilkan "please fill out this field" pada input bagian kata kunci <i>Hasil yang diharapkan :</i>	Valid

			
	Mereset form yang telah terisi dengan menekan tombol "reset"	Sistem akan menghapus semua form yang telah terisi	
3	<i>Test Case :</i> 	<i>Hasil yang diharapkan :</i> 	Valid
	Mengisi form hanya bagian kata kunci dan langsung klik tombol "kirim"	Sistem akan memprosesnya dan menampilkan hasilnya	
4	<i>Test Case :</i> 	<i>Hasil yang diharapkan :</i> 	Valid
	Mengisi semua form dan klik tombol "kirim"	Sistem akan memproses masukan dari pengguna	
5	<i>Test Case :</i> 	<i>Hasil yang diharapkan :</i> 	Valid

V. KESIMPULAN

Setelah merancang sistem analisis sentimen yang dibuat dalam sebuah aplikasi web, maka penulis dapat mengambil kesimpulan sebagai berikut :

1. Sistem analisis opini ini menyediakan form masukan (input) salah satunya untuk memasukan kata kunci yang ingin dicari arah sentimennya sesuai dengan keinginan pengguna.
2. Sistem opini ini memakai data tweets atau cuitan yang sesuai dengan kata kunci yang di masukan oleh pengguna dan secara khusus hanya akan mencari cuitan berbahasa Indonesia.
3. Dalam mengklasifikasikan cuitan kedalam 3 jenis opini atau sentimen yaitu positif, negatif dan netral, sistem ini memakai metode Klasifikasi Naive Bayes yang telah terbukti bisa mengklasifikasi data text seperti cuitan dari pengguna Twitter dengan menggunakan data training berupa kamus bahasa Inggris yang diterjemahkan kedalam bahasa Indonesia.

REFERENSI

- [1] Sunni, Ismail, and Dwi Hendratmo Widyantoro. 2012. Analisis sentimen dan ekstraksi topik penentu sentimen pada opini terhadap tokoh publik. ISSN: 2302-0040. Bandung: Jurnal Sarjana ITB bidang Teknik Elektro dan Informatika, Vol. 1, No. 2, Juli 2012: 200-206. Diambil dari : <https://www.academia.edu/download/30162931/156-609-1-PB.pdf> (16 Juni 2020)
- [2] Putranti, Novia Dwi, and Edi Winarko. 2014. Analisis sentimen twitter untuk teks berbahasa Indonesia dengan maximum entropy dan support vector machine. ISSN: 2460-7258. Yogyakarta: *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)* Vol. 8, No. 1 Januari 2014: 91-100. Diambil dari : <https://dev.jurnal.ugm.ac.id/ijccs/article/view/3499/3025> (16 Juni 2020)
- [3] Hadna, Nurrun Muchammad Shiddieqy, Paulus Insap Santosa, and Wing Wahyu Winarno. 2016. Studi literatur tentang perbandingan metode untuk proses analisis sentimen di Twitter. ISSN: 2089-9815. Yogyakarta: Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016). Maret 2016: 57-64. Diambil dari : <https://fti.uajy.ac.id/sentika/publikasi/makalah/2016/95.pdf> (14 Juni 2020)
- [4] Guntoro, Ghulam Asrofi. 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. ISSN: 2477-5274. Ponorogo: *INTEGER: Journal of Information Technology*, Vol.2, No. 1, Maret 2017: 32-41. Diambil dari : <https://ejurnal.itats.ac.id/integer/article/view/95/136> (16 Juni 2020)
- [5] Hidayatullah, Ahmad Fathan, and Azhari SN Azhari. 2015. Analisis sentimen dan klasifikasi kategori terhadap tokoh publik pada twitter. ISSN: 1979-2328. Yogyakarta: Seminar Nasional Informatika (SEMNASIF). Vol. 1. No. 1. Agustus 2015: 115-122. Diambil dari : <http://www.jurnal.upnyk.ac.id/index.php/semnasif/article/download/1016/977> (14 Juni 2020)
- [6] Priyanti, Dwi, and Siska Iriani. 2013. Sistem Informasi Data Penduduk Pada Desa Bogoharjo Kecamatan Ngadirojo Kabupaten Pacitan. ISSN: 2354-6654. Surakarta: *IJNS - Indonesian Journal on Networking and Security* Vol. 2, No. 4 Oktober 2013: 55-61. Diambil dari : <http://ijns.org/journal/index.php/ijns/article/view/181/176> (16 Juni 2020)
- [7] Gunawan, Billy, Helen Sasty Pratiwi, and Enda Yudha Pratama. 2018. Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. ISSN: 2548-9364. Pontianak: *JEPIN (Jurnal Edukasi dan Penelitian Informatika)* Vol. 4, No. 2 Desember 2018: 113-118. Diambil dari : <https://core.ac.uk/download/pdf/294889001.pdf> (14 Juni 2020)
- [8] Sipayung, Evasaria Magdalena, Herastia Maharani, and Ivan Zefanya. 2016. Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. ISSN: 2355-4614. Palembang: *Jurnal Sistem Informasi*, Vol. 8, No. 1, April 2016: 958-965. Diambil dari : <https://core.ac.uk/download/pdf/267823675.pdf> (16 Juni 2020)
- [9] Febriani, Mega. 2014. Strategi Komunikasi Pemasaran dalam Membangun Brand Image pada Media Sosial Twitter @Pekanbaruco. ISSN: 2355-6919. Pekanbaru: Jurnal Online Mahasiswa Fakultas Ilmu Sosial dan Ilmu Politik Universitas Riau, Vol. 1, No. 2, Oktober 2014: 1-15. Diambil dari: <https://www.neliti.com/publications/31419/strategi-komunikasi-pemasaran-dalam-membangun-brand-image-pada-media-sosial-twit#cite> (14 Juni 2020)
- [10] Annur, Haditsah. 2018. Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. ISSN: 2087-1716. Makassar: *ILKOM Jurnal Ilmiah* Vol. 10, No. 2 Agustus 2018: 160-165, Diambil dari : <http://jurnal.fikom.umi.ac.id/index.php/ILKOM/article/view/303/142> (14 Mei 2020)
- [11] Saleh, Alfa. Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. ISSN: 2354-5771. Medan; *Creative information technology (Citec) Journal* Vol. 2, No. 3, Mei 2015:207-217. Diambil dari : <https://citec.amikom.ac.id/main/index.php/citec/issue/view/12> (16 Juni 2020)



Apif Supriadi. Lahir di Kuningan, 27 Mei 1998. Tahun 2020 lulus dari Program Strata Satu (S1) Jurusan Sistem Informasi di STMIK Antar Bangsa. Saat ini aktif sebagai Anggota Bidang Hubungan Masyarakat Gerakan Arah Baru Indonesia (GARBI) Kabupaten Tangerang.



Fatmasari, M.Kom. Lahir pada Tanggal 08 April 1978. Tahun 2006 lulus dari Program Strata Satu (S1) Proram Studi Fakultas Teknologi Informasi / Sistem Informasi. Tahun 2010 lulus dari Program Magister (S2) Program Studi Magister Ilmu Komputer / Teknik Informatika. Tahun 2018 mengikuti Kegiatan Profesional / Pengabdian Kepada Masyarakat dengan jenis kegiatan Pemanfaatan Teknologi Edukasi Smartphone Untuk Mendukung Pembelajaran Anak. Saat ini aktif sebagai Dosen Tetap di STMIK Antar Bangsa. Aktif melakukan penelitian yang dipublikasikan pada Jurnal Ilmiah. Diantaranya : Aplikasi Kriptografi SMS dengan Memanfaatkan AES 256, Blowfish dan DES Berbasis Anroid, Implementasi Algoritma Dijkstra untuk menentukan Jarak Terdekat Pencarian outlet Minimal Store berbasis Android pada Area Jabodetabek.