

## Analisis Sentimen Maskapai Citilink Pada Twitter Dengan Metode Naïve Bayes

Moh.Yasid<sup>a</sup>, Lukman Junaedi<sup>b</sup>

<sup>a</sup>Program Studi Teknik Informatika Universitas Narotama, Jl. Arief Rachman Hakim No.51, Surabaya, 60117, Indonesia

<sup>b</sup>Akademi Sistem Informasi, Universitas Narotama, Jl. Arief Rachman Hakim No.51, Surabaya, 60117, Indonesia

### INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 16 Agustus 2019

Revisi Akhir: 09 September 2019

Diterbitkan Online: 30 September 2019

### KATA KUNCI

Analisis

Sentimen

Maskapai

Citilink

Naïve Bayes

### KORESPONDENSI

No HP: +6281334052655

E-mail: [jsyasid@gmail.com](mailto:jsyasid@gmail.com)

### A B S T R A C T

*The use of airline transportation is no longer a new thing for the community, ranging from business people, families going on vacation, even to the reach of the middle class. This transportation is really needed by the community to be able to travel long distances with a relatively shorter period of time. One of them is Citilink, Citilink is a subsidiary airline of the Garuda Indonesia company with ticket prices that can be reached by the middle class and below. Assessment of Citilink airlines can be seen through any promos that are most promoted to the public. The good and bad services of Citilink Airlines as an evaluation for people who want to use these services, can also be seen from the tweet of the people on Twitter social media. Therefore many researchers use social media as a place to collect data. This research aims to classify or group positive or negative sentiments of the community towards Citilink airline services in order to provide information about public satisfaction with Citilink Indonesia airline services, with a dataset of 2000 tweets resulting in accuracy reaching 0.778, with a processing time of two minutes twenty three seconds.*

## 1. PENDAHULUAN

Berkembangnya industri jasa penerbangan yang semakin meningkat di Indonesia pada setiap tahunnya telah menjadi perhatian masyarakat luas. Maskapai penerbangan yakni suatu perusahaan atau sarana transportasi di butuhkan banyak masyarakat agar dapat bepergian jauh dengan jangka waktu yang relatif lebih singkat. Salah satunya adalah Citilink, citilink merupakan maskapai penerbangan anak dari perusahaan Garuda Indonesia dengan harga tiket yang dapat dijangkau masyarakat menengah kebawah. Penilaian terhadap maskapai penerbangan Citilink bisa di lihat melalui promo-promo apa saja yang paling banyak di promosikan kepada masyarakat, namun banyak juga masyarakat memberikan penilaian melalui pendapat atau komentar yang dikemukakan di media sosial Twitter. Twitter merupakan satu dari banyak medial sosial yang diminati banyak masyarakat. Pendapat tersebut berupa tweet yang nantinya bisa tersebar di timeline Twitter.

Banyaknya pengguna Twitter memberikan pengaruh besar bagi masyarakat dalam memberikan penilaian berupa pendapat atau komentar terhadap jasa maskapai penerbangan Citilink karena perkembangan pengguna Twitter yang semakin meningkat setiap tahunnya. Hal ini di kemukakan oleh Dick Costolo, CEO Twitter, ketika melaksanakan kunjungan di Indonesia dengan jumlah pengguna sudah mencapai 50 juta di Indonesia, menurut situs [tekno.kompas.com](http://tekno.kompas.com) (2015). Pendapat

atau komentar di media sosial twitter tersebut dapat dipakai pula sebagai data penelitian analisis sentimen untuk mengelompokan atau mengklasifikasi sentimen positif atau negatif.[1]

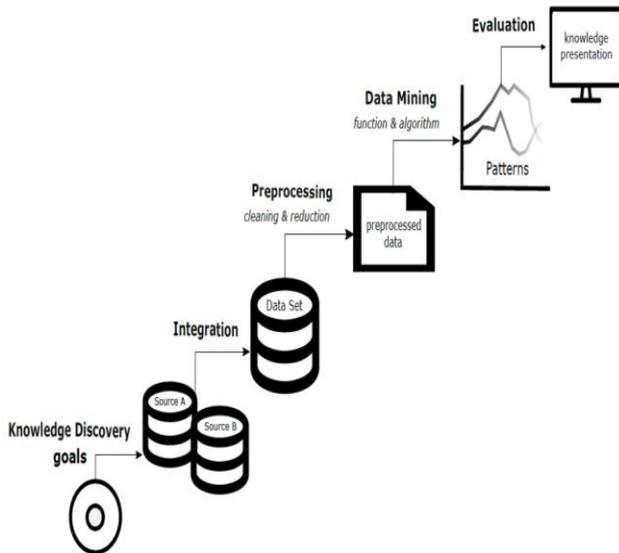
Analisis sentimen adalah suatu proses pengolahan kata untuk mengelompokan atau mengklasifikasi pendapat, penilaian, seseorang terkait kreasi, organisasi, atau aktivitas tertentu.[2] Penelitian kali ini menggunakan Naïve Bayes bertujuan melakukan klasifikasi atau pengelompokan sentimen positif atau negatif masyarakat terhadap layanan maskapai penerbangan Citilink berbahasa Indonesia agar bisa memberikan informasi tentang kepuasan masyarakat terhadap layanan jasa maskapai penerbangan Citilink Indonesia.

## 2. TINJAUAN PUSTAKA

Analisis sentimen atau opinion mining yakni proses memahami, mengekstrak dan mengolah data tekstual untuk memperoleh informasi yang terdapat dalam suatu kalimat pendapat atau opini.[3], [4]. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap suatu masalah atau objek oleh seseorang, apakah cenderung punya pemikiran atau beropini positif atau negatif.

Data *mining* adalah suatu proses untuk mengolah data mentah menjadi informasi yang bermanfaat. Saat ini data mining dipakai hampir pada semua aspek yang melibatkan data-data

digital.[5] Hal ini yang kemudian menimbulkan kekhawatiran terkait keamanan dan etika di dalam pelaksanaannya, karena istilah data mining sendiri dapat di artikan penggalian data, dan sering juga dikaitkan dengan penggalian data terkait perilaku manusia. Berikut sedikit gambaran tentang tahapan pada data mining.



Gambar 1. Tahapan Data Mining

Data Mining yakni tahapan yang memakai ilmu matematika, ilmu statistika serta kecerdasan buatan atau *machine learning* untuk melakukan ekstrasi data serta melakukan pengidentifikasian terhadap informasi yang terkandung dalam data yang diolah. Dalam data mining terdapat beberapa istilah yang biasa digunakan seperti KDD (*Knowledge discovery in data base*), analisa bentuk data atau (*Data Analysis*), ekstraksi pemahaman (*Knowledge Extraction*), Kecerdasan bisnis (*Business Intellegence*), *data archeology*, *data dregging*. [6]

Text Mining adalah salah satu proses yang bisa dipakai untuk mengerjakan klasifikasi dokumen yang berusaha untuk menemukan bentuk yang menarik dari sekumpulan data yang terkumpul.[7]

*Text mining* merupakan bagian dari data mining dimana proses yang dilakukan utamanya adalah mengerjakan ekstraksi pengetahuan dan informasi dari pola-pola yang terdapat dalam sekumpulan dokumen teks memakai alat analisis tertentu. [8] Text Mining yakni analisis kata, data diperoleh dari dokumen yang bermaksud untuk mencari kata yang dapat menggantikan dari dokumen kata-kata yang lain.[5], [9].

Twitter adalah salah satu layanan media sosial yang cukup terkenal dan memungkinkan para penggunanya untuk menulis status yang sering dinamakan kicauan atau tweet. Media sosial Twitter digunakan untuk mengungkapkan berbagai pendapat atau opini akan suatu produk, layanan atau hal lainnya.

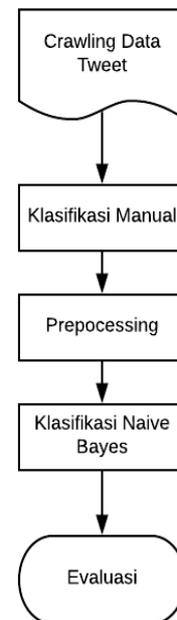
Twitter diciptakan oleh Jack Dorsey di tahun 2006 dan pertama meluncur di dunia maya saat Juli 2006 dengan alamat <http://www.Twitter.com> yang masih digunakan hingga saat ini. Pengguna dapat menulis pesan berdasarkan topik dengan menggunakan tanda #(hashtag). Sedangkan untuk menyebutkan atau membalas pesan dari pengguna lain bisa menggunakan tanda @.

Naive Bayes (Naive Bayes Classifier) adalah algoritma yang sangat efektif dalam permasalahan klasifikasi atau penggolongan. Algoritma ini bekerja berdasarkan probabilitas yang sudah ada untuk menentukan probabilitas yang akan datang. Meskipun dataset yang digunakan ada jutaan jumlahnya, Naive Bayes adalah pendekatan yang dianjurkan untuk digunakan. Untuk memahami lebih jauh tentang Naive Bayes maka perlu terlebih dahulu memahami dalil Bayes. Dalil Bayes dinamakan sesuai dengan nama penemunya Rev. Thomas Bayes. [10] Algoritma ini bekerja berdasarkan probabilitas bersyarat, yaitu probabilitas yang menjelaskan tentang sesuatu akan terjadi berdasarkan peristiwa yang telah terjadi sebelumnya.

Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes dengan asumsi independensi (ketidak tergantungan) yang kuat atau naif.[11]

Metode Bayesian classification digunakan menganalisis dalam membantu tercapainya pengambilan keputusan terbaik suatu permasalahan dari sejumlah alternatif. Kaitan antara Naive Bayes dengan klasifikasi, korelasi hipotesis, dan bukti dengan klasifikasi adalah hipotesis dalam teorema Bayes merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur fitur yang menjadi masukan dalam model klasifikasi.[6]

### 3. METODOLOGI



Gambar 2. Diagram Alur Penelitian

Diatas adalah tahapan-tahapan untuk menggambarkan proses klasifikasi dari awal sampai akhir dengan menggunakan metode Neive Bayes. Keluaran dari penelitian ini merupakan data klasifikasi berupa pendapat positif atau negatif dan besarnya akurasi yang akan diperoleh dari metode Naive bayes. Penelitian diawali dari proses memasukkan (input) data tweet hasil crawling. Dokumen hasil crawling akan diklasifikasi manual dengan memberikan identitas (label) positif atau negatif pada dataset, dan evaluasi sebagai tahap akhir.

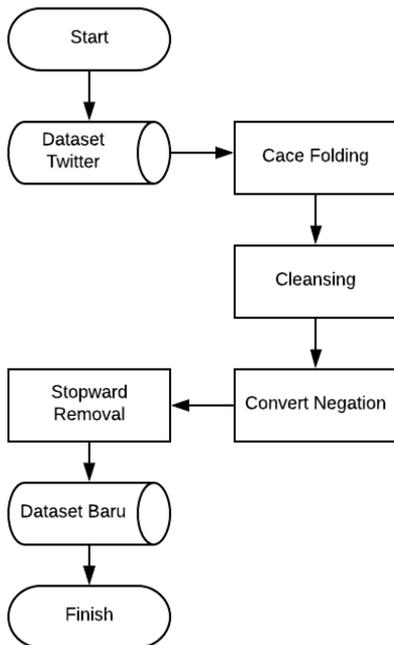
#### 4. HASIL DAN PEMBAHASAN

##### Pengumpulan Dataset

Data didapatkan dari hasil proses Crawling data tweet. Crawling merupakan proses pengumpulan data dari media sosial twitter dengan melakukan pencarian melalui suatu kata kunci (keywords) tertentu. Pada proses crawling ini menggunakan library twitter scrapper dengan kata kunci @citolink, citilink. Data akan diklasifikasi secara manual dengan memberikan label positif atau negatif. Klasifikasi label positif dan negatif pada dataset dilakukan secara manual dengan bantuan kata-kata besar (big og words) sebagai kamus, yang akan dijadikan input pada penelitian ini. Berikut contoh dokumen input terdapat pada tabel 1.

Table 1. Contoh Dokumen Input

Opini	Tweet
Positive	@Citilink #Terima kasih Citilink atas pelayanan prima merubah last min tiket mudik kt krn wafatnya ayah kami td mlm.Semoga makin sukses
Negative	@Citilink pramugari tidak ramah, hanya berdiri saja tidak membantu penumpang



Gambar 3. Tahapan Preprocessing

Diatas adalah gambaran alur tahap-tahap dari proses *preprocessing*. Berikut penjelasan pada setiap tahap-tahap diatas :

##### Case Folding

Tahap *case folding* merupakan proses penyamaan huruf pada dokumen tweet dengan mengubah huruf kapital jadi huruf kecil, supaya merenggangkan kata berlebih-lebihan dikarenakan perbedaan huruf, dan akan menghasilkan dokumen tweet baru. Berikut contoh tweet yang dihasilkan dari tahap case folding bisa dilihat pada tabel 2.

Table 2. Tahap Case Folding

Input	Output
@Citilink #Terima kasih Citilink atas pelayanan prima merubah last min tiket mudik kt krn wafatnya ayah kami td mlm.Semoga makin sukses	@citilink #terima kasih citilink atas pelayanan prima merubah last min tiket mudik kt krn wafatnya ayah kami td mlm.Semoga makin sukses

##### Cleansing

Tahap *cleansing* adalah proses yang bertujuan untuk menghilangkan tweet dari kata yang kurang dibutuhkan seperti simbol HTML, kata kunci, hastags(#), username, email dan simbol (!@#%<>[]=;:','^\$\*+\_-/?), yang akan menghasilkan dokumen tweet baru. Berikut contoh dokumen tweet yang menjadi keluaran dari tahap cleansing bisa dilihat pada tabel 3.

Table 3. Tahap Cleansing

Input	Output
@Citilink #Terima kasih Citilink atas pelayanan prima merubah last min tiket mudik kt krn tiket mudik kt krn wafatnya ayah kami td ayah kami td mlm.Semoga makin sukses, '.	terima kasih Citilink atas pelayanan prima merubah last min tiket mudik kt krn wafatnya ayah kami td mlm.Semoga makin sukses,

##### Convert Negation

Proses *convert negation* bertujuan untuk transformasi kata negasi yang tergabung dalam tweet. Kata tersebut akan mengganti arti suatu sentimen pada tweet tersebut, kata negasi akan tergabung dengan kata setelahnya. Contoh kata negasi 'tidak', banyak lagi kata negasi yang akan menghasilkan dokumen tweet baru. Berikut contoh tweet yang dihasilkan dari tahap convert negation terdapat pada tabel 4.

Table 4. Tahap Convert Negation

Input	Output
citilink pramugari tidak ramah, hanya berdiri saja tidak membantu penumpang	citilink pramugari tidakramah, hanya berdiri saja tidak membantu penumpang

##### Stopword Removal

Stopword Removal, yakni teknik yang bertujuan untuk menghapus kata didalam dokumen dataset, kata tersebut seperti "di", "ke", "dari", "yang", "sedang", "ini", karena

dianggap tidak memberikan deskripsi sesuatu dalam bahasa Indonesia. Tapi pada teks klasifikasi kata semacam “tidak”, “bukan”, “tanpa” kurang penting sehingga kata tersebut tidak ikut dibuang. Berikut contoh hasil tahap stopword removal terdapat pada tabel 5.

**Table 5.** Tahap Stopword Removal

Input			Output		
citilink	terima	kasih	Citilink	terima	kasih
Citilink	atas	pelayanan	pelayanan	prima	merubah
prima	merubah	last min	last min	tiket	mudik kt
tiket	mudik kt	krn wafatnya	wafatnya	ayah kami	mlm
ayah kami	td	mlm.Semoga	Semoga	sukses	
makin	sukses				

**Klasifikasi Naïve Bayes**

Pada penelitian klasifikasi dengan metode Naïve Bayes diperlukan nilai (bobot) pada setiap kata dalam dokumen dataset yang akan diproses. Nilai (bobot) diperoleh dengan menghitung peluang (probabilitas) positif atau negatif terhadap kata dalam setiap dokumen *dataset*.

Penghitungan peluang (probabilitas) positif bertujuan untuk menghitung nilai (bobot) pada tiap kata dalam dokumen dataset. Data yang diperlukan dalam penghitungan ini berupa banyaknya kata berkategori positif, kemunculan tiap kata dalam dokumen yang berkategori positif, dan total keseluruhan kata. Nilai tersebut nantinya akan dipakai pada klasifikasi naïve bayes, akan berpengaruh dalam menentukan besarnya akurasi yang menjadi parameter kebenaran pada penelitian ini.

Penghitungan peluang (probabilitas) negatif bertujuan untuk menghitung nilai (bobot) pada tiap kata dalam dokumen dataset. Data yang diperlukan dalam penghitungan ini berupa banyaknya kata berkategori negatif, kemunculan tiap kata dalam dokumen yang berkategori negatif, dan total keseluruhan kata. Nilai tersebut nantinya akan dipakai pada klasifikasi naïve bayes akan berpengaruh dalam menentukan besarnya akurasi yang menjadi parameter kebenaran pada penelitian ini.

Klasifikasi sentimen dengan naïve bayes dilakukan dengan cara membandingkan nilai (bobot) kata pada dokumen dalam dataset. Bila nilai (bobot) kata peluang (probabilitas) yang berkategori positif lebih banyak maka hasil sentimennya positif, bila nilai (bobot) kata peluang (probabilitas) yang berkategori negatif maka hasil sentimennya negatif Semua dokumen dataset akan diklasifikasi bila ditemukan nilai (bobot) pada tiap kata di dokumen dataset, dan data tidak terklasifikasi bila tidak ditemukan nilai (bobot) pada tiap kata di dokumen dataset.

Dari hasil pengujian dataset dengan jumlah 2000 tweet, terdapat 1661 dokumen tweet terklasifikasi, dan 339 tweet tidak terklasifikasi, seperti pada tabel 6 berikut:

**Table 6.** Data Klasifikasi

Dokumen Tweet	Classified	Unclassified
2000	1661	339

**Evaluasi**

Evaluasi bertujuan untuk mengecek kebenaran hasil klasifikasi dengan menghitung berapa angka yang diperoleh. Acuan yang dipakai untuk tahap evaluasi yakni perhitungan tabel matriks klasifikasi (*confussion matrix*). Berikut tabel perhitungan *confussion matrix* bisa dilihat pada tabel 7.

**Table 7.** Confussion Matrix

		True Class	
		Positive	Negative
Kelas Terprediksi	Positif	TP(True Positif)	FP(False Positive)
	Negatif	FN(False Negative)	TN(True Negative)

Dari dokumen *tweet* sebanyak 2000 yang telah disebutkan diatas, menghasilkan perhitungan *confussion matrix* sebagai berikut.

**Confussion Matrix**

Actual / Predicted	Positive	Negative	Total
Positive	1286	35	1321
Negative	333	7	340
Total	1619	42	1661

**Gambar 4.** Penghiungan Confussion Matrix

Dari hasil perhitungan *confussion matrix* diatas, dimana data *tweet* sudah terklasifikasi, akan dihitung untuk melihat akurasi yang akan didapat sebagai acuan dari kebenaran klasifikasi pada penelitian ini. Berikut hasil perhitungan *confussion matrix* terdapat pada gambar 3.

**Hasil Evaluasi**

Evaluasi	Nilai
<b>Akurasi</b>	0.77844671884407
<b>Recall Positive</b>	0.97350492051476
<b>Precision Positive</b>	0.79431747992588
<b>Recall Negative</b>	0.020588235294118
<b>Precision Negative</b>	0.166666666666667

**Gambar 5.** Hasil Evaluasi

Evaluasi dipakai sebagai parameter dari kebenaran dalam penelitian ini, dengan akurasi perhitungan dari tabel *confussion matrix* (matriks klasifikasi). Untuk menghitung akurasi, *precision*, *recall*. Akurasi yang didapat mencapai 0.778 dengan waktu proses dua menit dua puluh tiga detik. Nilai akurasi adalah tingkat keberhasilan klasifikasi dari semua data. *Precision* adalah data actual positif yang diprediksi benar. *Recall* untuk mengukur keberhasilan klasifikasi yang diprediksi dengan benar.

## 5. KESIMPULAN DAN SARAN

Setelah dilaksanakannya penelitian dengan tahap-tahap dapat disimpulkan bahwa klasifikasi dokumen tweet berbahasa Indonesia menggunakan metode Naïve Bayes untuk melihat penilaian masyarakat terhadap layanan maskapai penerbangan Citilink Indonesia melalui media sosial twitter dengan dataset sebanyak 2000 tweet menghasilkan akurasi yang mencapai 0,778, waktu proses dua menit dua puluh tiga detik dengan kode pemrograman PHP. Dan juga menambahkan kosa kata bahasa Inggris yang sering digunakan masyarakat dalam memberikan opini mengenai maskapai penerbangan pada kamus, seperti *delay, flight, boarding, landing, reschedule, onschedule, cancel, thanks*.

untuk peneliti selanjutnya penulis menyarankan untuk menambahkan tahap seleksi fitur seperti contoh: seleksi fitur Information Gain, atau yang lainnya, menambahkan semua kosa kata bahasa Inggris karena dimungkinkan warga Negara asing juga memakan layanan maskapai penerbangan yang diharapkan akan memperbaiki nilai akurasi.

## DAFTAR PUSTAKA

- [1] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel," *J. Khatulistiwa Inform.*, vol. VI, no. 1, 2018.
- [2] "Text Mining Dan Sentimen Analisis Twitter Pada Gerakan Lgbt," *Intuisi J. Psikol. Ilm.*, vol. 9, no. 1, pp. 18–25, 2017.
- [3] V. Effendy, "ANALISIS SENTIMEN BERBAHASA INDONESIA DENGAN PENDEKATAN LEXICON BASED ( STUDI KASUS : SOLUSI PENGELOLAAN SAMPAH ) Jurnal Ilmiah Komputer dan Informatika ( KOMPUTA )," vol. 4, no. 1, 2015.
- [4] I. Zulfa and E. Winarko, "Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 11, no. 2, p. 187, 2017.
- [5] W. Gata, "Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS," vol. 6, pp. 1–13, 2017.
- [6] D. Wahyudi, A. H. Mirza, and P. H. Merriayu, "IMPLEMENTASI DATA MINING DENGAN NAIVE BAYES CLASSIFIER UNTUK Mendukung Strategi Promosi ( Studi kasus Universitas Bina Darma Palembang )," pp. 1–10.
- [7] N. MOH, "Klasifikasi Dokumen Komentar Pada Situs Youtube Menggunakan Algoritma K-Nearest Neighbor (K-Nn)," *Univ. Dian Nuswantoro*, no. 5, 2016.
- [8] R. Delima, U. Kristen, D. Wacana, A. Rachmat, U. Kristen, and D. Wacana, "Implementasi Metode K-Nearest Neighbor dengan Decision Rule untuk Klasifikasi Subtopik Berita IMPLEMENTASI METODE K-NEAREST NEIGHBOR DENGAN DECISION RULE UNTUK KLASIFIKASI per topik-topik berita sehingga membuat pencarian berita dipermudah . Bagi Pembuat," no. June, 2014.
- [9] K. Teks and B. Indonesia, "Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks Bahasa," no. May, 2016.
- [10] A. Rachmat C and Y. Lukito, "Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes," *J. Inform. dan Sist. Inf. Univ. Ciputra*, vol. 02, no. 02, pp. 26–34, 2016.
- [11] R. W. Pratiwi and Y. S. Nugroho, "Prediksi Rating Film Menggunakan Metode Naïve Bayes," *J. Tek. Elektro*, vol. 8, no. 2, pp. 60–63, 2016.

## BIODATA PENULIS



### Moh. Yasid

Mahasiswa Program Studi Teknik Informatika, Fakultas Ilmu Komputer, universitas Narotama Surabaya  
Email : [jsyasid@gmail.com](mailto:jsyasid@gmail.com)



### Lukman Junaedi

Dosen di Departemen Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi, Universitas Narotama, Surabaya, Indonesia. Ia memperoleh gelar masternya M.Kom. dari Institut Teknologi Sepuluh Nopember (ITS), Surabaya Indonesia pada 2012.  
Email : [lukman.junaedi@narotama.ac.id](mailto:lukman.junaedi@narotama.ac.id)