



**Analisa Opini Terhadap Produk Perangkat Keras IT
Pada Media Sosial Dengan Klasifikasi Menggunakan
Algoritma Naive Bayes Dan Lexicon Based**

TUGAS AKHIR

Aditya Pratama Yudha
41516120022

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2020**



**Analisa Opini Terhadap Produk Perangkat Keras IT
Pada Media Sosial Dengan Klasifikasi Menggunakan
Algoritma Naive Bayes Dan Lexicon Based**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Aditya Pratama Yudha
41516120022

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2020

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41516120022

Nama : Aditya Pratama Yudha

Judul Tugas Akhir : Analisa Opini Terhadap Produk Perangkat Keras IT Pada Media Sosial Dengan Klasifikasi Menggunakan Algoritma Naive Bayes Dan Lexicon Based

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 27 November 2020



Aditya Pratama Yudha

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Aditya Pratama Yudha
NIM : 41516120022
Judul Tugas Akhir : Analisa Opini Terhadap Produk Perangkat Keras IT Pada Media Sosial Dengan Klasifikasi Menggunakan Algoritma Naive Bayes Dan Lexicon Based

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

UNIVERSITAS
MERCU BUANA

Jakarta, 27 November 2020



Aditya Pratama Yudha

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Aditya Pratama Yudha
 NIM : 41516120022
 Judul Tugas Akhir : Analisa Opini Terhadap Produk Perangkat Keras IT Pada Media Sosial Dengan Klasifikasi Menggunakan Algoritma Naive Bayes Dan Lexicon Based

Menyatakan bahwa Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	
		Jurnal Nasional Terakreditasi	✓
		Jurnal International Tidak Bereputasi	
		Jurnal International Bereputasi	Diterima
Disubmit/dipublikasikan di :	Nama Jurnal	: Analisa Opini Terhadap Produk Perangkat Keras IT Pada Media Sosial Dengan Klasifikasi Menggunakan Algoritma Naive Bayes Dan Lexicon Based	
	ISSN	: 2579-8901	
2	Kertas Kerja, Merupakan material hasil penelitian sebagai kelengkapan Artikel Jurnal. Terdiri dari (minimal 4)	Literatur Review	[✓]
		Hasil analisa & perancangan aplikasi	[✓]
		Source code	[✓]
		Data set	[✓]
		Tahapan eksperimen	[✓]
		Hasil eksperimen seluruhnya	[✓]
		
3	Disubmit / Terdaftar	HKI	Diajukan
		Paten	Tercatat
		No & Tanggal Permohonan	
		No & Tanggal Pencatatan	

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 15 April 2021



Aditya Pratama Yudha

LEMBAR PERSETUJUAN PENGUJI

NIM : 41516120022
Nama : Aditya Pratama Yudha
Judul Tugas Akhir : Analisa Opini Terhadap Produk Perangkat Keras IT Pada Media Sosial Dengan Klasifikasi Menggunakan Algoritma Naive Bayes Dan Lexicon Based

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 8 Februari 2021



(Desi Ramayanti, S.Kom., MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41516120022
Nama : Aditya Pratama Yudha
Judul Tugas Akhir : Analisa Opini Terhadap Produk Perangkat Keras IT Pada Media Sosial Dengan Klasifikasi Menggunakan Algoritma Naive Bayes Dan Lexicon Based

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 8 Februari 2021



(Afiyati, S.Si, MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41516120022
Nama : Aditya Pratama Yudha
Judul Tugas Akhir : Analisa Opini Terhadap Produk Perangkat Keras IT Pada Media Sosial Dengan Klasifikasi Menggunakan Algoritma Naive Bayes Dan Lexicon Based

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 8 Februari 2021




LEMBAR PENGESAHAN

NIM : 41516120022
Nama : Aditya Pratama Yudha
Judul Tugas Akhir : Analisa Opini Terhadap Produk Perangkat Keras IT Pada Media Sosial Dengan Klasifikasi Menggunakan Algoritma Naive Bayes Dan Lexicon Based

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.


Jakarta, 8 Februari 2021

Menyetujui,


(Dr. Leonard Goeirianto)
Dosen Pembimbing

UNIVERSITAS
MERCU BUANA

Mengetahui,


(Diky Firdaus, S.Kom, MM)

Koord. Tugas Akhir Teknik Informatika


(Desi Ramayanti, S.Kom, MT)

Ka. Prodi Teknik Informatika

ABSTRAK

Nama : Aditya Pratama Yudha
NIM : 41516120022
Pembimbing TA : Leonard Goeirmanto S.T, Dr., M.Sc
Judul : Analisa Opini Terhadap Produk Perangkat Keras IT Pada Media Sosial Dengan Klasifikasi Menggunakan Algoritma Naive Bayes Dan Lexicon Based

Analisis ini dilakukan untuk mendapatkan opini publik terhadap suatu produk, yang nantinya informasinya dapat digunakan dalam penjualan produk tersebut. Apakah produk tersebut memiliki reputasi yang baik dan bagaimana opini publik terhadap produk sejenis tetapi memiliki merek yang berbeda akan dibahas lebih lanjut dalam tulisan ini. Dalam makalah ini diperlukan data yang diperoleh dari media sosial seperti Twitter, dimana data mentahnya akan diolah terlebih dahulu kemudian dilakukan proses analisis opini dengan menggunakan dua algoritma yang berbeda yaitu algoritma Naive Bayes untuk prediksi dan Algoritme berbasis leksikon untuk pemrosesan awal data. Penelitian ini dilakukan karena beberapa orang masih bingung dalam memilih produk dengan tipe yang sama namun dengan merk yang berbeda khususnya VGA card untuk penelitian ini. Untuk itu, hasil penelitian ini diharapkan dapat bermanfaat bagi masyarakat yang ingin mencari / membeli produk agar tidak salah dalam memilih produk tersebut. Berdasarkan penelitian sebelumnya terkait dengan analisis opini yang membedakan hanya algoritma yang digunakan dan dataset yang digunakan juga berbeda namun tetap tujuannya adalah analisis opini juga. Yang membedakan penelitian ini dengan penelitian sebelumnya, pada penelitian sebelumnya hanya proses identifikasi topik tertentu saja, sedangkan penelitian ini lebih menitikberatkan pada produk IT khususnya VGA Card. Dan metode serta algoritma yang digunakan juga berbeda-beda, kali ini penulis mencoba melakukan metode klasifikasi dengan menggunakan algoritma Naive Bayes dengan beberapa algoritma modifikasi dan algoritma berbasis leksikon untuk preprocessing data. Penelitian ini akan dilakukan dalam beberapa tahapan yaitu proses pengumpulan data, proses pembersihan data atau tahap preprocessing, proses penerapan algoritma terkait, dan menampilkan hasil dengan visualisasi data dengan wordcloud untuk setiap topik dan hasil prediksi algoritma.

Kata kunci:

penelitian, analisa opini, sentiment analisis, klasifikasi, algortima lexicon based, algortima naïve bayes classifier

ABSTRACT

Name : Aditya Pratama Yudha
Student Number : 41516120022
Counsellor : Leonard Goeirmanto S.T, Dr., M.Sc
Title : Opinion Analysis Of IT Hardware Products On
Social Media With Classification Using The Naive
Bayes Algorithm And Lexicon Based

This analysis is carried out to obtain public opinion on a product, which the information can later be used in selling the product. Whether the product has a good reputation and how public opinion on similar products but has a different brand will be discussed further in this paper. In this paper, data obtained from social media such as Twitter, for example, is needed, which the raw data will be processed first and then the opinion analysis process will be carried out using two different algorithms, they are Naive Bayes algorithm for predicting and Lexicon-based algorithm for data preprocessing. This research was conducted because some people are still confused in choosing a product with the same type but with a different brand especially the VGA card for this research . For this reason, the results of this research are expected to be useful for people who want to find / buy products so they are not wrong for choosing those products. Based on previous research related to opinion analysis which distinguishes only the algorithms used and the dataset used is also different but still the goal is opinion analysis as well. What distinguishes this research from previous research, in previous studies only the identification process of a particular topic, while this research focuses more on a IT product especially the VGA card. And the methods and algorithms used are also different, at this time the author tries to perform a classification method using the Naive Bayes algorithm with some modification algorithm and lexicon based algorithm for data preprocessing. This research will be carried out in several stages, which is the data collection process, the data cleansing process or the preprocessing stage, the process of applying the related algorithms, and displaying the results with data visualisation with wordcloud for each topics and the results of the algorithm prediction.

Key words:

research, opinion analysis, sentiment analysis, classification, lexicon based algorithm, naïve bayes classifier algorithm.

KATA PENGANTAR

Puji syukur kita panjatkan kehadirat Allah SWT atas segala limpahan rahmat, taufik serta hidayah-Nya sehingga laporan tugas akhir ini dapat diselesaikan dengan waktu yang tepat dan juga cukup sesuai dengan apa yang diharapkan penulis. Penulis menyadari bahwa tanpa bantuan serta bimbingan dan arahan dari para dosen pembimbing dan juga dosen pengajar yang telah memberikan ilmu nya kepada penulis. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Leonard Goeirmanto S.T, Dr., M.Sc. selaku dosen pembimbing tugas akhir.
2. Ibu Dr. Devi Fitriana, S.Kom, MTI selaku dosen pembimbing akademik.
3. Para dosen Fakultas Ilmu Komputer Teknik Informatika terutama dosen yang telah mengajar penulis.
4. Orang tua penulis terutama Ibu yang selalu mensupport, memberikan semangat serta Doa.
5. Saudara kandung, teman serta rekan yang penulis kenal selama belajar di lingkungan kampus Universitas Mercubuana.

Namun demikian tidak tertutup kemungkinan masih adanya beberapa kekurangan pada laporan ini. Oleh karena itu, segala saran dan masukan dari semua pihak selalu diharapkan untuk perbaikan dan penyempurnaannya.

Akhir kata, penulis berharap laporan tugas akhir ini dapat berguna dan bermanfaat dalam menambah ilmu dan pengetahuan baru terhadap para pembaca laporan ini. Dan sekali lagi kepada semua pihak yang ikut berpartisipasi demi terwujudnya laporan tugas akhir ini penulis ucapkan terima kasih.

Jakarta, 5 Januari 2021

Aditya Pratam Yudha

DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii	
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI	v
LEMBAR PERSETUJUAN PENGUJI	vi
LEMBAR PERSETUJUAN PENGUJI	vii
LEMBAR PENGESAHAN	viii
ABSTRAK	ix
ABSTRACT.....	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xii
NASKAH JURNAL	1
KERTAS KERJA.....	10
BAB 1. LITERATUR REVIEW.....	11
BAB 2 ANALISIS DAN PERANCANGAN.....	15
BAB 3 SOURCE CODE	17
BAB 4 DATASET.....	25
BAB 5 TAHAPAN EKSPERIMEN	26
BAB 6 HASIL SEMUA EKSPERIMEN.....	80
BAB 7 HASIL REVISI EKSPERIMEN	87
DAFTAR PUSTAKA	107
LAMPIRAN TANDA TERIMA PENYERAHAN TUGAS AKHIR TU FASILKOM.....	108
LAMPIRAN DOKUMEN HAKI.....	109
LAMPIRAN KORESPONDENSI	111

xii

NASKAH JURNAL

Analisa Opini Terhadap Produk Perangkat Keras IT Pada Media Sosial Dengan Klasifikasi Menggunakan Algoritma Naive Bayes Dan Lexicon Based

Aditya Pratama Yudha¹, Leonard Goeirmanto²

*Program Studi Teknik Informatika, Fakultas Ilmu Komputer,
Universitas Mercubuana^{1,2}*

*41516120022@student.mercubuana.ac.id¹,
tugas@dataku.web.id²*

Abstrak - Analisis ini dilakukan untuk mendapatkan opini dari masyarakat terhadap suatu produk, yang informasinya nanti bisa digunakan dalam penjualan produk tersebut. Apakah produk tersebut memiliki reputasi yang baik di masyarakat dan bagaimana opini masyarakat terhadap produk yang sejenis namun memiliki brand yang berbeda akan dibahas lebih lanjut pada penelitian ini. Dalam penelitian ini dibutuhkan data yang didapat dari media sosial seperti twitter contohnya, yang kemudian data mentah tersebut akan diolah terlebih dahulu lalu kemudian akan dilakukan proses analisa opini menggunakan dua algoritma yang berbeda yaitu naive bayes dan lexicon based untuk melihat algoritma mana yang lebih efektif di antara keduanya. Penelitian ini dilakukan karena beberapa masyarakat masih bingung dalam memilih suatu produk dengan jenis yang sama namun dengan brand yang berbeda. Untuk itu dengan hasil dari penelitian ini diharapkan bermanfaat untuk masyarakat yang hendak mencari/membeli produk agar tidak salah dalam memilih produk tersebut. Berdasarkan penelitian sebelumnya terkait juga dengan analisa opini yang membedakan hanya algoritma yang digunakan saja serta dataset yang digunakan pun juga berbeda namun tetap tujuannya adalah analisa opini juga. Yang membedakan penelitian ini dengan penelitian sebelumnya, pada penelitian sebelumnya hanya proses identifikasi terhadap suatu topik tertentu sedangkan penelitian ini lebih berfokus kepada suatu produk. Dan metode serta algoritma yang digunakan pun juga berbeda, pada kali ini penulis mencoba melakukan metode klasifikasi dengan algoritma naive bayes serta lexicon based. Penelitian ini akan dilakukan dengan beberapa tahapan yaitu, proses pengumpulan data, proses pembersihan data (data cleansing) atau tahap preprocessing, proses penerapan algoritma terkait, serta menampilkan hasilnya dalam bentuk visualisasi data dan hasil prediksi algoritmanya.

Kata kunci: penelitian, analisa opini, sentiment analisis, klasifikasi, lexicon based, naive bayes classifier

1. PENDAHULUAN

I.I. Latar Belakang

Dewasa ini perkembangan data sangatlah pesat, karena segala sesuatu hal pasti memiliki data. Untuk itu penerapan data mining sangat berguna pada saat ini. Salah satu penerapan data mining pada penelitian ini adalah melakukan analisa opini dari media sosial terhadap suatu produk. Hasil dari analisa opini tersebut dapat berguna untuk calon pelanggan ke depan nya sebelum membeli barang yang akan dikehendaki apakah barang yang hendak mereka beli sudah tepat atau belum. Penelitian ini termasuk ke dalam *product reviews and feedback*. Untuk itu dengan hasil dari penelitian ini diharapkan bermanfaat untuk masyarakat yang hendak mencari/membeli produk agar tidak salah dalam memilih produk tersebut. Berdasarkan penelitian sebelumnya terkait juga dengan analisa opini yang membedakan hanya algoritma yang digunakan saja serta dataset yang digunakan pun juga berbeda namun tetap tujuannya adalah analisa opini juga. Yang membedakan penelitian ini dengan penelitian sebelumnya, pada penelitian sebelumnya hanya proses identifikasi terhadap suatu topik tertentu sedangkan penelitian ini lebih berfokus kepada suatu produk. Dan metode serta algoritma yang digunakan pun juga berbeda, pada kali ini penulis mencoba melakukan metode klasifikasi dengan algoritma naive bayes serta lexicon based. Penelitian ini akan dilakukan dengan beberapa tahapan yaitu, proses pengumpulan data, proses pembersihan data (data cleansing) atau tahap preprocessing, proses penerapan algoritma terkait, serta menampilkan hasilnya dalam bentuk visualisasi data dan hasil prediksi algoritma nya.

I.II. Rumusan Masalah

Penelitian ini dilakukan karena beberapa masyarakat masih bingung dalam memilih suatu produk dengan jenis yang sama namun dengan brand yang berbeda. Untuk itu dengan hasil dari penelitian ini diharapkan bermanfaat untuk masyarakat yang hendak mencari/membeli produk agar tidak salah dalam memilih produk tersebut. Berdasarkan penelitian sebelumnya terkait juga dengan analisa opini yang membedakan hanya Algoritma yang digunakan saja serta dataset yang digunakan pun juga berbeda namun tetap tujuannya adalah analisa opini juga.

I.III. Tujuan Penelitian

Tujuan dilaksanakannya penelitian ini yaitu membantu masyarakat memilih produk yang sesuai dengan keinginan mereka. Dan diharapkan masyarakat terbantu agar tidak salah dalam membeli/memilih suatu produk.

Manfaat yang didapat dari penelitian ini masyarakat mendapatkan informasi yang sesuai dan tepat mengenai produk yang hendak mereka cari/beli atas dasar opini orang lain yang beredar pada media sosial.

I.IV. Batasan Masalah

Batasan masalah yang akan dibahas pada penelitian ini yaitu:

- Penelitian hanya sebatas mencari opini masyarakat terhadap suatu produk.
- Penelitian bertujuan mencari hasil akurasi algoritma yang terbaik yang digunakan pada saat penelitian.
- Untuk jenis produk yang dibahas pada penelitian ini hanya seputar perangkat keras komputer / hardware.

- Untuk perangkat keras komputer / hardware yang menjadi batasan dalam topik penelitian ini yaitu adalah vga card (perangkat kartu grafis).

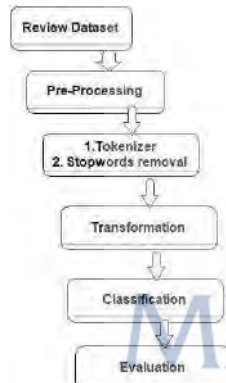
I.V. Landasan Teori

Berikut ini dasar teori yang terdapat pada penelitian :

Menentukan polaritas positif atau negatifnya suatu opini dapat dilakukan secara manual, tetapi seiring bertambahnya sumber opini menjadi semakin banyak tentunya waktu dan usaha yang dibutuhkan untuk mengklasifikasikan polaritas opini tersebut akan semakin banyak terpakai. Oleh karena itu, diajukan penerapan metode *machine learning* untuk mengklasifikasi polaritas opini dari sumber data yang sangat banyak tersebut. Untuk melakukan hal itu, bisa menggunakan salah satu fungsi dari text mining, dalam hal ini adalah klasifikasi dokumen. (Nurhuda, Sihwi, Doewes 2013).

Penelitian lain dilakukan oleh Felida (2017) menggunakan Naïve Bayes untuk memprediksi Waktu Produksi Furniture di Gajendra Furniture dengan tujuan menghasilkan akurasi produksi sebesar 79%. Penelitian lebih lanjut diharapkan dapat mengoptimalkan untuk menghasilkan akurasi yang lebih tinggi terhadap jenis produk yang berbeda dalam hal ini adalah *vga card*.

I.VI. Literatur Review



Gambar 1 Diagram Proses Opinion Mining

Review Dataset

Adalah tahap pemilihan data dan crawling data dan melihat apakah data tersebut bisa diolah atau tidak.

Pre-Processing

Adalah tahap pengolahan data dari raw data (data mentah) ke data yang clean dan sudah bisa diolah menggunakan algoritma terkait.

Tokenizer/Tokenisasi

Tokenisasi adalah proses untuk membagi teks yang dapat berupa kalimat, paragraf atau dokumen, menjadi token-token/bagian-bagian tertentu.

Stopwords Removal

Stopword removal disebut juga filtering, adalah tahap pemilihan kata-kata penting dari hasil token, yaitu kata-kata apa saja yang akan digunakan untuk mewakili dokumen. Proses stemming adalah proses pembentukan kata dasar.

Classification

Classification memiliki tujuan untuk mengklasifikasikan suatu data ke dalam kelompok kelas yang sudah ada. Tidak akan ada pembentukan kelompok baru. Dan biasanya prosesnya supervised. Dalam arti ada training yang dilakukan. Salah satu pendekatan classification adalah pendekatan neural network. Dalam pendekatan ini, neural network dilatih dengan menggunakan data yang sudah diketahui untuk dimasukkan ke dalam kelompok data tertentu. Setelah pelatihan, prosesnya cukup sederhana, yaitu dengan memasukkan data baru (yang tidak termasuk dalam data training) ke dalam neural network, lalu neural network akan memberikan informasi termasuk ke dalam kelompok manakah data baru tersebut.

Evaluation

evaluasi adalah kegiatan yang dilakukan berkenaan dengan proses untuk menentukan nilai dari suatu hal.

METODOLOGI PENELITIAN

1. Jenis Data

Dataset yang digunakan pada penelitian ini adalah data tweet pengguna twitter mengenai suatu produk. Dalam hal ini adalah produk IT yaitu VGA Card memiliki brand yang berbeda seperti VGA Card RADEON dan VGA Card NVIDIA karena kedua produk tersebut memiliki fungsi yang sama namun dari brand yang berbeda.

Dataset diperoleh melalui *crawling* data menggunakan Twitter API. Yang memiliki jumlah 3500 record pada masing-masing file. Dan untuk kolom nya berjumlah 16 kolom.

2. Metode Pengumpulan data

Karena topik penelitian ini adalah data mining jadi metode pengumpulan data yang digunakan untuk penelitian adalah crawling data dari media sosial seperti twitter atau media sosial lain nya. Data yang diperoleh masih dalam bentuk raw data / data mentah yang masih harus diolah lagi atau dilakukan tahap pre-processing dataset agar menjadi data yang siap untuk di analisa. Dalam hal ini penulis menggunakan Twitter API key yang berisi *consumer key*, *consumer secret*, *access token* dan *access secret*. Kemudian lakukan pencarian seperti memasukkan kata kunci yang diinginkan, menggunakan library *rtweet/twitteR* Dan save hasilnya ke dalam bentuk raw data (rds format).

3. Model Matematika

Misalkan S menjadi sistem yang menjelaskan ekstraksi tweet, Preprocessing, Pelabelan sentimen, Analisis Sentimen-

$S = \{Tw, Pt, Sl\}$

Tw = Tweet diekstrak dari Twitter.

Sl = {Pw, Nw}

Pw = {P1, P2, ..., Pn} = Kelas Positif

Nw = {N1, N2, ..., Nn} = Kelas Negatif

Dimana,

S = Sistem analisa sentiment.

Pt = Pra-pemrosesan Tweet (Terjemahan kata gaul, Penghapusan kata non-Inggris, penandaan PoS, Penghapusan URL dan pemberhentian kata).

Sl = Sentiment Labeling menggunakan Sent Strength dan Twitter sentiment analysis tools (Naïve Bayes untuk memberikan hasil yang lebih efisien dan akurat).

P1, p2...Pn kelas koleksi tweet positif

N1, N2...Nn Kelas koleksi tweet negatif

4. Algoritma Matematika

Merujuk pada jurnal (Wagh, Shinde, Wankhade 2016), Teorema Bayes memberikan cara untuk menghitung probabilitas posterior, $P(c | x)$, dari $P(c)$, $P(x)$, dan $P(x | c)$. Pengklasifikasi Naive Bayes mengasumsikan bahwa pengaruh nilai prediktor (x) pada kelas tertentu (c) tidak bergantung pada nilai prediktor lain. Asumsi ini disebut kebebasan bersyarat kelas.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

The diagram shows the formula $P(c|x) = \frac{P(x|c)P(c)}{P(x)}$ with arrows pointing to each part: 'Likelihood' points to $P(x|c)$, 'Class Prior Probability' points to $P(c)$, 'Posterior Probability' points to $P(c|x)$, and 'Predictor Prior Probability' points to $P(x)$.

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Gambar 2 Rumus Matematika

Penjelasan rumus diatas,

$P(c | x)$ adalah probabilitas posterior kelas (c , target) yang diberi prediktor (x , atribut).

$P(c)$ adalah probabilitas kelas sebelumnya.

$P(x | c)$ adalah kemungkinan yang merupakan probabilitas kelas prediktor yang diberikan.

$P(x)$ adalah probabilitas prediktor sebelumnya.

5. Metode Penelitian

Tahapan selanjutnya adalah *pre-processing* dan *cleansing dataset* menggunakan library tm. Dan untuk fitur yang digunakan seperti menghapus url pada dataset. menghapus garis baru/new lines, menghapus tanda baca koma, menghapus kata RT (retweet), menghapus tanda baca titik dua, menghapus tanda baca titik koma, menghapus tanda baca titik, menghapus special character/symbol, menghapus tanda @ dan nama user, serta menghapus spasi yang berlebihan.

Kemudian melakukan proses *stopwords removal* dalam hal ini metode yang digunakan adalah *lexicon based* dengan menggunakan kamus kata/*word dictionary* sendiri, dan agar hasilnya lebih optimal penulis juga menggunakan fitur *dictionary stopwords(english)* dari library *tm*.

Langkah selanjutnya adalah membuat *document term-matrix (DTM)*, DTM berfungsi sebagai matriks matematika yang menggambarkan frekuensi istilah yang terjadi dalam kumpulan dokumen. Dan melakukan visualisasi data seperti membuat *word cloud* dari dataset yang sudah bersih.



Gambar 3 word cloud dataset radeon



Gambar 4 word cloud dataset nvidia

Proses selanjutnya adalah melakukan *labeling dataset* dan juga proses *sentiment analysis*, load dataset yang sudah bersih selanjutnya tambahkan *word dictionary* untuk kata positif dan juga negatif lalu tambahkan beberapa kata selain dari *word dictionary* agar hasil *labeling* nya lebih optimal. Terakhir masukkan dan jalankan *function* untuk melakukan proses *labeling* dan *sentiment analysis* pada dataset nya. Proses pelabelan data tergantung dari dataset yang di load semakin besar dataset nya maka proses nya akan memakan waktu yang cukup lama juga. Hasil dari *labeling* ini tipe nya berupa *scoring* yaitu score 0 (neutral), -1 s.d -5 (negative) dan 1 s.d 5 (positive). Lalu ubah hasil *scoring* ke bentuk kata positive atau negative agar selanjutnya dataset dapat diolah menggunakan algoritma *naïve bayes classifier*.

	classification	score	text
1	Positive	0	whats radeon
2	Positive	0	weve rthnered uwith give werful gaming pc
3	Positive	0	weve rthnered uwith give werful gaming pc
4	Positive	0	nice exriment amd radeon handle strife rtxue nvtbcausti...
5	Positive	1	pretty ebay ad images xt sale lol
6	Negative	-2	shit year radeon graics card run cold war ony
7	Positive	0	mega teaming scan gift pc featurng br
8	Positive	0	check dell sehgx inch full hd gaming monitor hz tn ms ...
9	Positive	0	indiedevhour comrison raytracing shadows left rrtfbre...
10	Positive	0	mega teaming scan gift pc featurng br

Gambar 5 Tampilan dataset yang sudah di labeling

6. Pengujian

Pada percobaan pertama penulis menggunakan algoritma naïve bayes classifier yang standar dan belum dimodifikasi. Setelah load dataset yang sudah di labeling lalu ubah ke bentuk *corpus object* dan select hanya text field lalu tentukan set seed dan bagi dataset menjadi training data dan juga testing data, kemudian lakukan juga pemisahan pada DTM nya menjadi training data dan juga testing data. Dan terakhir lakukan perhitungan *confusion matrix* setelah itu hasilnya akan terlihat untuk nilai skor akurasi prediksi nya.

Pada percobaan pertama hasil akurasi prediksi nya masih belum sesuai dengan harapan. Lalu kemudian penulis mencoba untuk melakukan modifikasi pada algoritma naïve bayes classifier nya. Dengan melakukan beberapa perubahan pada Algoritma nya pada percobaan kedua hasil akurasi nya meningkat dan bisa dikatakan cukup bagus untuk dataset Nvidia.

7. Evaluasi

Proses modifikasi algoritma sebenarnya belum maksimal masih bisa ditingkatkan lagi hasil nya namun karena keterbatasan waktu maka penulis belum mendapatkan hasil yang terbaik dari modifikasi algoritma naïve bayes classifier nya. Dan juga hasil dari proses preprocessing dan cleansing dataset ternyata cukup berpengaruh terhadap langkah selanjutnya yaitu pemrosesan algoritma naïve bayes classifier.

HASIL DAN PEMBAHASAN

Berikut adalah tabel perbandingan tingkat akurasi dari hasil pengujian nya:

Dataset	Skor Akurasi Prediksi Algoritma			
	Naïve bayes normal	Naïve bayes modifikasi		
		P1	P2	P3
Radeon	71,80%	74%	70,68%	72,28%
Nvidia	57,80%	87,70%	70,05%	74,28%

P1 : Percobaan pertama, rasio data traning:testing (50:50), modifikasi hanya pada set.seed dan dtm saja.

P2 : Percobaan kedua, rasio data training:testing (70:30), penyesuaian modifikasi pada algoritma pada bagian nrow, dtm dan lainnya.

P3 : Percobaan ketiga, rasio data training:testing (80:20), penyesuaian modifikasi pada algoritma pada bagian nrow, dtm dan lainnya.

Dari hasil pengujian diatas dapat disimpulkan bahwa tahap modifikasi algoritma ternyata cukup berpengaruh terhadap hasil skor akurasi prediksi. Dengan melakukan modifikasi yang tepat maka hasil yang didapat juga akan berdampak lebih baik pada skor akurasi prediksinya.

PENUTUP

1. Kesimpulan

Penelitian ini selesai sesuai dengan harapan penulis yaitu membandingkan dua jenis produk yang memiliki fungsi sama namun dengan brand/merk yang berbeda. Dalam penelitian ini dataset yang digunakan adalah dari opini publik di social media twitter terhadap suatu produk dalam hal ini perangkat keras komputer yaitu vga card. Dengan adanya penelitian ini diharapkan calon konsumen lebih yakin dalam membeli barang yang diinginkan terutama perangkat keras komputer terutama vga card berdasarkan dari opini publik di social media.

2. Saran

Penulis sangat berterima kasih dan berkenan menerima masukan pembaca sekalian jika memiliki saran yang membangun untuk penelitian ini agar penelitian ini menjadi lebih baik lagi kedepannya.

DAFTAR PUSTAKA

(Hanggara et al., 2017; Hegde et al., 2018; Le & Nguyen, 2015; Martínez-Cámara et al., 2014; Meduru et al., 2017; Moore, n.d.; Nurhuda & Sihwi, 2014; Of et al., n.d.; Perdana & Pinandito, 2018; Thesis, 2012; Wagh et al., 2016; Zhang & Gao, 2011)

- [1] Hanggara, S., Akhriza, T. M., & Husni, M. (2017). Aplikasi Web Untuk Analisis Sentimen Pada Opini Produk Dengan Metode Naive Bayes Classifier. *Seminar Nasional Inovasi Dan Aplikasi Teknologi Di Industri*, 4(2), A33.1-A33.6.
- [2] Hegde, B., S, N. H., & Prakash, M. (2018). Sentiment analysis of Twitter data: A machine learning approach to analyse demonetization tweets. *International Research Journal of Engineering and Technology*. www.irjet.net
- [3] Le, B., & Nguyen, H. (2015). 3rd International Conference on Computer Science, Applied Mathematics and Applications, ICCSAMA 2015. *Advances in Intelligent Systems and Computing*, 358, 1–415. <https://doi.org/10.1007/978-3-319-17996-4>
- [4] Martínez-Cámara, E., Martín-Valdivia, M. T., Ureña-López, L. A., & Montejor-Ráez, A. R. (2014). Sentiment analysis in Twitter. *Natural Language*

- Engineering*, 20(1), 1–28. <https://doi.org/10.1017/S1351324912000332>
- [5] Meduru, M., Mahimkar, A., Subramanian, K., Padiya, P. Y., & Gunjgur, P. N. (2017). Opinion Mining Using Twitter Feeds for Political Analysis. *International Journal of Computer (IJC)*, 25(1), 116–123. <http://ijcjournal.org/>
- [6] Moore, A. W. (n.d.). *slides - Naive Bayes Classifiers*.
- [7] Nurhuda, F., & Sihwi, S. W. (2014). Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier. *ITSmart: Jurnal Ilmiah Teknologi Dan Informasi*, 2(2), 35–42.
- [8] Of, I., Bayes, N., For, C. A., & Sofa, I. O. N. (n.d.). *DETERMINING CUSTOMERS ' BUYING*. 127–134.
- [9] Perdana, R. S., & Pinandito, A. (2018). Combining likes-retweet analysis and naive bayes classifier within twitter for sentiment analysis. *Journal of Telecommunication, Electronic and Computer Engineering*, 10(1–8), 41–46.
- [10] Thesis, B. (2012). *Erik Lux Feature selection for text classification with Naive Bayes*.
- [11] Wagh, B., J. V., S., & N. R., W. (2016). Sentimental Analysis on Twitter Data using Naive Bayes. *Ijarcce*, 5(12), 316–319. <https://doi.org/10.17148/ijarcce.2016.51273>
- [12] Zhang, W., & Gao, F. (2011). An improvement to naive bayes for text classification. *Procedia Engineering*, 15, 2160–2164. <https://doi.org/10.1016/j.proeng.2011.08.404>



KERTAS KERJA

Ringkasan

Bagian ini merupakan ringkasan dari lembar kerja, untuk dataset yang digunakan terdapat pada bagian 4 sedangkan penjelasan dan penggunaan source code nya terdapat pada bagian 3. Dan berikut adalah hasil eksperimen secara keseluruhan.

Dari hasil pengujian skor akurasi prediksi algoritma untuk dataset *radeon* menggunakan algoritma naïve bayes normal mendapatkan hasil 71,80% dan setelah menggunakan algoritma algoritma naïve bayes modifikasi hasilnya naik menjadi 74%. Sedangkan untuk dataset *nvidia* menggunakan algoritma naïve bayes normal mendapatkan hasil 57,80% dan setelah menggunakan algoritma algoritma naïve bayes modifikasi hasilnya naik menjadi 87,70%. Dari percobaan tersebut dapat disimpulkan bahwa tahap modifikasi algoritma ternyata cukup berpengaruh terhadap hasil skor akurasi prediksi. Dengan melakukan modifikasi yang tepat maka hasil yang di dapat juga akan berdampak lebih baik pada skor akurasi prediksi nya.



UNIVERSITAS
MERCU BUANA