

Analisis Penentuan Respons Twitter sebagai Media Komunikasi dan Informasi Pemerintah Berbasis Metode Rabin-Karp

Luh Ayu Fernita¹, N. P. Sastra², Rukmi Sari Hartati³

[Submission: 23-07-2023, Accepted: 14-09-2023]

Abstract— Social media in this digital era, especially Twitter, has an important role in the interaction between government and society. Given the wide variety of activities within government circles, the task of monitoring and responding to messages is complex and time-consuming. Therefore, this study aims to develop an effective approach in determining the appropriate response from the government to people's tweets. This study proposes using a combination of the Rabin-Karp method to quickly determine relevant responses. The Rabin-Karp method, known for its efficiency in pattern matching, is used to match tweets to a set of tweets that have already been given a response. Furthermore, the Word2Vec technique is used to improve understanding of the meaning of the text. The use of the Rabin-Karp method with the addition of the Word2Vec method shows that the response accuracy rate is 74.55%. The results of this study also show that the lower the K-Gram value, the higher the similarity value, and vice versa. These results are expected to contribute in the context of government that is responsive to societal issues discussed on Twitter.

Intisari— Media sosial dalam era digital ini, khususnya Twitter, memiliki peran penting dalam interaksi antara pemerintah dan masyarakat. Adanya beragam kegiatan di lingkungan pemerintahan menimbulkan tugas memantau dan menanggapi pesan menjadi rumit dan memakan waktu. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan pendekatan yang efektif dalam menentukan respons yang tepat dari pemerintah terhadap tweet masyarakat. Penelitian ini, mengusulkan penggunaan kombinasi metode Rabin-Karp untuk menentukan respons yang relevan dengan cepat. Metode Rabin-Karp, yang dikenal karena efisiensinya dalam pencocokan pola, digunakan untuk mencocokkan tweet dengan kumpulan tweet yang sudah diberikan respons. Selanjutnya, digunakan teknik Word2Vec untuk meningkatkan pemahaman makna teks. Penggunaan metode Rabin-Karp dengan penambahan metode Word2Vec menunjukkan tingkat akurasi respons adalah 74,55 %. Hasil penelitian ini juga menunjukkan bahwa semakin rendah nilai K-Gram maka semakin tinggi nilai similaritasnya, begitu pula sebaliknya. Hasil ini diharapkan memberikan kontribusi dalam konteks pemerintahan yang responsif terhadap isu-isu masyarakat yang dibahas di Twitter.

¹Program Magister Teknik Elektro, Br. Batanbuah, Ds.Tangguntiti, Kec. Selemadeg Timur, Tabanan, 82162 (telp:087852034112; e-mail: ayudiahfernita@gmail.com)

^{2,3} Program Pasca Sarjana, Magister Teknik Elektro Fakultas Teknik Universitas Udayana, Jln. P.B. Sudirman, Denpasar, Bali 80232 INDONESIA (telp: 0361-703315; fax: 0361-4321; e-mail: ²putra.sastra@unud.ac.id, ³rukmisari@unud.ac.id,

penelitian ini juga menunjukkan bahwa nilai K-Gram yang semakin rendah, menyebabkan nilai similarity akan semakin tinggi, begitu juga sebaliknya.

Kata Kunci— Rabin-Karp; Word2Vec; Twitter, Respons

I. PENDAHULUAN

Perkembangan pesat internet telah mengubah lanskap media sosial dengan cepat. Media sosial memberikan wadah online yang memudahkan pengguna dalam hal komunikasi dan informasi [1][2]. Salah satu media sosial yang populer adalah Twitter yang telah menjadi sarana penting untuk menyampaikan aspirasi dan berkomunikasi antara masyarakat dan pemerintah [3]. Dalam konteks pelayanan publik, keterlibatan aktif masyarakat dan respons cepat dari pemerintah sangat penting untuk meningkatkan kualitas layanan [4]. Namun, terdapat tantangan dalam mengelola dan merespons volume besar tweet yang masuk ke akun pemerintah. Oleh karena itu, diperlukan metode yang efektif untuk mengidentifikasi dan menentukan respons yang sesuai terhadap tweet yang berkaitan dengan pemerintah.

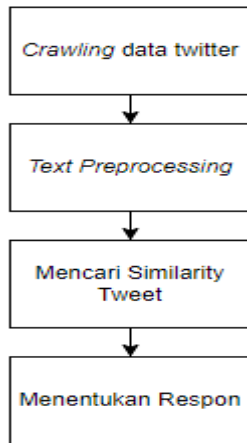
Penelitian ini bertujuan untuk mengatasi tantangan tersebut dengan memanfaatkan metode Rabin-Karp dalam analisis teks tweet. Metode Rabin-Karp adalah algoritma pencarian string yang menggunakan fungsi hashing untuk mendeteksi pola dalam teks. Keunggulan utama metode ini adalah Rabin-Karp dapat mencocokkan lebih dari satu pola secara efisien dengan menggunakan pendekatan rolling hash [5]. Hal ini memungkinkan pencarian simultan untuk beberapa pola dalam teks yang sama. Dalam penelitian ini, metode Rabin-Karp akan diterapkan untuk menemukan kesamaan teks pada media Twitter, dengan fokus pada tweet yang terkait dengan akun pemerintah.

Studi kasus dalam penelitian ini adalah Pemerintahan Kabupaten Badung, yang telah menjadi salah satu kota menuju smart city. Kabupaten Badung memiliki kebutuhan yang tinggi dalam merespons aspirasi masyarakat melalui media sosial, terutama Twitter. Dengan menerapkan metode Rabin-Karp, diharapkan dapat ditemukan tweet dengan kesamaan teks tertinggi, yang kemudian akan dijadikan acuan untuk menentukan respons yang tepat dari akun pemerintah. Penelitian ini memiliki potensi untuk meningkatkan efisiensi dan efektivitas komunikasi antara masyarakat dan pemerintah melalui media Twitter, sehingga memberikan kontribusi positif dalam pelayanan publik dan pengambilan keputusan pemerintah yang.



II. DESAIN SISTEM

Penelitian ini terdiri dari empat tahap penelitian yang meliputi koleksi data, pra-pemrosesan teks, metode kesamaan, dan penentuan hasil. Rincian dari setiap tahapan tersebut dijelaskan secara lebih lanjut pada Gambar 1.



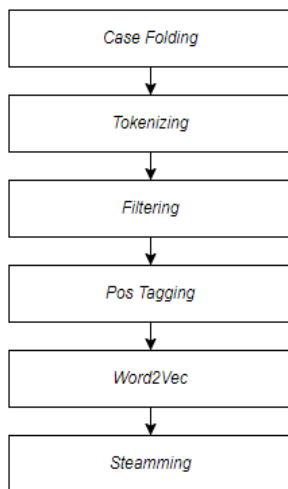
Gambar 1. Alur Penelitian

A. Koleksi Data

Data yang digunakan pada penelitian ini diperoleh dari twitter, lebih detailnya adalah data yang berhubungan dengan Pemerintah Kabupaten Badung. Dari kumpulan data tersebut akan dibagi menjadi data latih dan data uji. Data yang didapatkan sudah diberikan respons dan kategori untuk masing-masing tweet. Adapun kategori yang diberikan dibagi menjadi 3 yaitu informasi, pertanyaan dan pengaduan

B. Text Preprocessing

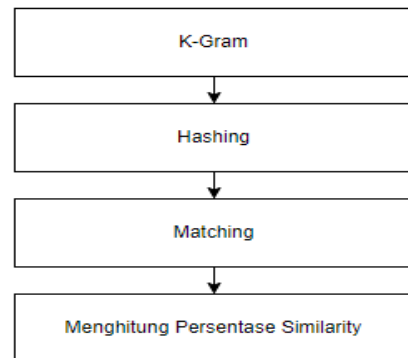
Text preprocessing merupakan proses yang dilakukan untuk memperoleh informasi yang berkualitas tinggi dari kumpulan teks [6][7]. Adapun tahapan yang dilakukan pada proses ini dijelaskan pada Gambar 2 [8].



Gambar 2. Tahapan Text Processing

C. Metode Similarity

Tahap berikutnya adalah *pencarian similarity* data twitter antara data uji dan data latih. Metode *similarity* yang digunakan adalah metode Rabin-Karp. Proses-proses yang dilakukan pada tahapan ini dijelaskan pada Gambar 3.



Gambar 3. Tahapan Metode Rabin Karp

D. Penentuan Hasil Similarity

Penentuan hasil dilakukan setelah mendapatkan persentase *similarity* dengan masing-masing data tweet dari twitter pada data latih. Selanjutnya akan dicari persentase tertinggi dan jika nilai tersebut lebih dari 80%, proses dilanjutkan dengan menentukan jenis respons twitter. Sebaliknya jika nilai kurang dari 80%, data akan ditambahkan ke daftar *uncategorized* untuk selanjutnya data tersebut direspons secara manual. Tipe tweet dalam penelitian ini dibagi menjadi tiga, yaitu pertanyaan, pengaduan, dan informasi. Setiap tweet akan diberikan respons sesuai dengan tipe tweet dengan kemiripan tertinggi.

III. HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh melalui proses pengumpulan yang hati-hati dari tiga sumber utama: Twitter, Aplikasi Sidumas, dan melalui konsultasi dengan Bidang Pengelolaan Informasi Publik (PIP) di Pemerintahan Kabupaten Badung. Langkah-langkah yang dilakukan dalam pengumpulan data ini diuraikan sebagai berikut:

Pengumpulan Data dari Twitter (@PembkabBadung)

Data diperoleh dengan menggunakan antarmuka pemrograman aplikasi (API) yang disediakan oleh Twitter untuk mengambil tweet yang mencantumkan " @PembkabBadung" sebagai mention. Respons atau tanggapan yang diberikan oleh akun @PembkabBadung untuk masing-masing tweet diverifikasi untuk mendapatkan data yang akurat dan valid.

Pengumpulan Data dari Aplikasi Sidumas

Data yang berasal dari aplikasi Sidumas, platform pengaduan yang dimiliki oleh Pemerintah Badung, diperoleh melalui akses ke sistem tersebut. Data pengaduan akan dijadikan sebagai alternatif data tweet.

TABEL III
 CONTOH HASIL TOKENIZING

Kata Awal	Hasil Kata
lampu penerangan jalan padam dan sangat membahayakan pada malam hari mohon pihak terkait membantu	{lampu}, {penerangan}, {jalan}, {padam}, {dan}, {sangat}, {membahayakan}, {pada}, {malam}, {hari}, {mohon}, {pihak}, {terkait}, {membantu}

Konsultasi dengan Bidang PIP

Melalui pertemuan dan konsultasi dengan Bidang Pengelolaan Informasi Publik (PIP), informasi langsung diperoleh dari sumber yang memiliki wewenang di dalam Pemerintahan Kabupaten Badung. Pertanyaan-pertanyaan yang diajukan melalui konsultasi dari bidang tersebut digunakan sebagai sumber informasi yang sah. Data pertanyaan ini dijadikan sebagai data tambahan tweet

Validasi dan Pemberian Respon Data

Validasi lintas-sumber dilakukan dengan membandingkan dan memeriksa konsistensi data dari ketiga sumber. Data kemudian diklasifikasikan berdasarkan kategori yang telah ditentukan. Setiap data diberikan respon dan kategori yang disesuaikan isi tweet. Berdasarkan pencarian data yang dilakukan dari tahun 2021 dan 2022 didapatkan data sejumlah 485. Data tersebut akan dibagi menjadi 2 bagian, 30% digunakan sebagai data uji dan 70% digunakan sebagai data latih. Semua data tersebut diberikan respons yang sudah dikoordinasikan dengan pihak terkait sehingga setiap tweet memiliki jawaban masing-masing.

B. Hasil Text Preprocessing

Text preprocessing merupakan tahap pada pengolahan data yang memiliki tujuan untuk dapat membersihkan dan mempersiapkan data teks mentah (*raw text*) sebelum dilakukan analisis atau diolah lebih lanjut [9]. Adapun tahapan yang dilakukan dengan contoh salah satu data yang digunakan adalah sebagai berikut.

Case Folding

Tahapan ini adalah tahapan pengubahan karakter dalam teks menjadi huruf *lowercase* atau huruf kecil [10]. Contoh proses *case folding* ditunjukkan pada Tabel I.

TABEL I
 CONTOH HASIL CASE FOLDING

Kata Awal	Hasil Kata
lampu penerangan jalan padam dan sangat membahayakan pada malam hari Mohon pihak terkait membantu	lampu penerangan jalan padam dan sangat membahayakan pada malam hari mohon pihak terkait membantu

Tokenizing

Tokenizing adalah tahapan pemisahan kumpulan kata pada kalimat, paragraf atau halaman menjadi potongan kalimat tunggal atau token yang akan menjadi data kata yang berdiri sendiri [11]. Contoh hasil *tokenizing* ditunjukkan pada Tabel II.

Filtering

Tahapan ini menyeleksi kata-kata yang signifikan dari hasil proses tokenisasi, yaitu kata-kata yang dapat mewakili konten suatu dokumen[12]. Adapun contoh dari filtering ditunjukkan pada Tabel III.

TABEL IIIII
 CONTOH HASIL FILTERING

Kata Awal	Hasil Kata
{lampu}, {penerangan}, {jalan}, {padam}, {dan}, {sangat}, {membahayakan}, {pada}, {malam}, {hari}, {mohon}, {pihak}, {terkait}, {membantu}	{lampu}, {penerangan}, {jalan}, {padam}, {sangat}, {membahayakan}, {malam}, {hari}, {mohon}, {pihak}, {terkait}, {membantu}

Pos Tagging

Pemberian label kelas kata merupakan proses yang dilakukan pada tahapan *Pos Tagging*, hal ini akan dilakukan dengan otomatis terhadap suatu kata pada rangkaian kalimat [13]. Khusus untuk kata verb, pada penelitian ini akan dilakukan *Word2Vec* karena kata kerja mengandung informasi penting tentang tindakan atau kegiatan. Hasil proses ini ditunjukkan pada Tabel IV.

TABEL IVV
 CONTOH HASIL POS TAGGING

Kata	Pos Tagger
lampu	JJ
penerangan	NN
jalan	NN
padam	VB
sangat	RB
membahayakan	VB
malam	NN
hari	NN
mohon	NN
pihak	NN
terkait	VB
membantu	VB

Word2Vec

Word2vec merupakan pemodelan *Natural Language Processing* (NLP) yang digunakan untuk menghasilkan representasi vektor kata-kata dalam teks [14]. Tahapan *Word2Vec* bertujuan untuk menambah kata-kata yang memiliki makna serupa atau terkait [15]. Arsitektur yang digunakan pada metode ini adalah *Continuous Bag-of-Words*



(CBOW). Penelitian ini menggunakan Library Gensim, yaitu library dalam bahasa pemrograman Python yang digunakan untuk NLP dan pemodelan topik [16]. Contoh script yang digunakan adalah sebagai berikut.

```
from gensim.models import Word2Vec
sentences = Word2Vec.load('Model file Location')
model = Word2Vec(sentences, min_count=1)
word = {word}
similar_words = model.wv.most_similar(word)
```

Sedangkan contoh hasil dari Word2Vec ini ditunjukkan pada Tabel V.

TABEL V
CONTOH HASIL WORD2VEC

Kata Awal	Hasil Kata
padam	{mati},{tewas},{berhenti},{hilang}, {berhenti}
membahayakan	{'merugikan},{mengganggu}, {mengancam},{merusak}, {membebani},{mempersulit}, {berbahaya},{menyulitkan}, {mengganggu},{menggangu'}
terkait	{'berkaitan},{berkenaan}, {berhubungan},{sehubungan}, {kaitannya},{menyangkut}, {dikaitkan},{menggaitkan}, {bertalian'}
membantu	{'membantunya},{mendorong}, {menolong},{memotivasi}, {mengawasi},{berusaha}, {mempercepat},{mempermudah}, {berupaya},{memberdayakan}'}

Hasil dari Word2Vec akan ditambahkan kedalam *text preprocessing* untuk meningkatkan kualitas analisis teks dengan memungkinkan model memahami makna kata-kata dalam konteks yang lebih luas dan lebih mendalam. Adapun hasil *text preprocessing* setelah Word2Vec yaitu

```
{lampu}, {penerangan}, {jalan},
{mati},{tewas},{berhenti},{hilang}, {berhenti}, {sangat},
{membahayakan},{merugikan}, {mengganggu},
{mengancam}, {merusak}, {membebani}, {mempersulit},
{berbahaya}, {menyulitkan}, {mengganggu}, {menggangu'},
{malam}, {hari}, {mohon}, {pihak}, {terkait}'{berkaitan},
{berkenaan}, {berhubungan}, {sehubungan}, {kaitannya},
{menyangkut}, {dikaitkan}, {menggaitkan}, {bertalian'},
{membantu}, {'membantunya}, {mendorong}, {menolong},
{memotivasi}, {mengawasi}, {berusaha}, {mempercepat},
{mempermudah}, {berupaya}, {memberdayakan}'}
```

Stemming

Stemming merupakan tahapan untuk mengubah struktur kata ke bentuk dasar atau mengidentifikasi akar kata dari hasil kata setelah melalui proses refleksi [17]. Contoh hasil *stemming* ditunjukkan pada Tabel VI.

TABEL VI
CONTOH HASIL STEAMMING

Kata Awal	Kata Hasil	Kata Awal	Kata Hasil
lampu	lampu	pihak	pihak
penerangan	terang	berkaitan	kait
jalan	jalan	berkenaan	kena
mati	mati	berhubungan	hubung
tewas	tewas	sehubungan	hubung
berhenti	berhenti	kaitan	kait
hilang	hilang	menyangkut	sangkut
merugikan	rugi	dikaitkan	kait
mengganggu	ganggu	bertalian	tali
mengancam	ancam	membantunya	bantu
merusak	rusak	mendorong	dorong
membebani	beban	menolong	tolong
mempersulit	sulit	memotivasi	motivasi
berbahaya	bahaya	mengawasi	awas
menyulitkan	sulit	berusaha	usaha
mengganggu	ganggu	mempercepat	cepat
malam	malam	mempermudah	mudah
hari	hari	berupaya	upaya
mohon	mohon	memberdayakan	daya

C. Hasil Metode Similarity

Metode *similarity* Rabin-Karp digunakan untuk menentukan kemiripan text. Langkah-langkah yang dilakukan adalah sebagai berikut.

K-Gram

K-Gram adalah serangkaian istilah dengan panjang k. K-Gram digunakan untuk mengekstrak huruf-huruf karakter sejumlah k dari sebuah kata yang terus-menerus dibaca dari teks sumber hingga akhir dokumen [18]. Contoh penerapan K-Gram dengan nilai k = 5 adalah sebagai berikut.

- Kata hasil *text preprocessing*
{lampu}, {terang}, {jalan}, {mati}, {tewas}, {berhenti}, {hilang}, {rugi}, {ganggu}, {ancam}, {rusak}, {beban}, {sulit}, {bahaya}, {sulit}, {ganggu}, {malam}, {hari}, {mohon}, {pihak}, {kait}, {kena}, {hubung}, {hubung}, {kait}, {sangkut}, {kait}, {tali}, {bantu}, {dorong}, {tolong}, {motivasi}, {awas}, {usaha}, {cepat}, {mudah}, {upaya}, {daya}
- Kata digabungkan
lampu terang jalan mati tewas berhenti hilang rugi gangguan cam rusak bebansulit bahayasulit ganggumalam hari mohon pihakkaitkenahubunghubungkait sangkut kait talibantudorong tolong motivasi awas usahacepat mudah upayadaya
- Hasil kata dengan panjang K-Gram
{'lampu}, {ampu}, {mpute}, {puter}, {utera}, {teran}, {erang}, {rangj}, {angja}, {ngjal}, {gjala}, {jalan}, {alanm}, {lanma}, {anmat}, {nmati}, {matit}, {atite}, {titew}, {itewa}, {tewas}, {ewasb}, {wasbe}, {asber}, {sberh}, {berhe}, {erhen}, {rhen}, {henti}, {entih}, {ntihi}, {tihil}, {ihil}, {hilan}, {ilang}, {langr}, {angru}, {ngrug}, {grugi}, {rugig}, {ugiga}, {gigan}, {igang}, {gangg}, {anggu}, {nggua}, {gguan}, {gguan}, {guanc}, {uancam}, {ancam}, {ncamr}, {camru}, {amrus}, {mrusa}, {rusak}, {usakb}, {sakbe}, {akbeb}, {kbeba}, {bebans}, {ebans},

{bansu}, {ansul}, {nsuli}, {sulit}, {ulitb}, {litba}, {itbah}, {tbaha}, {bahay}, {ahaya}, {hayas}, {ayas}, {yasul}, {asuli}, {sulit}, {ulitg}, {litga}, {itgan}, {tgang}, {gangg}, {anggu}, {nggum}, {gguma}, {gumal}, {umala}, {malam}, {alah}, {lahar}, {ahari}, {harim}, {arimo}, {rimoh}, {imoho}, {mohon}, {ohonp}, {honpi}, {onpih}, {npika}, {pikai}, {ika}, {kaitk}, {aitke}, {itken}, {tkena}, {kenah}, {ena}, {nahub}, {ahubu}, {hubun}, {ubung}, {bungk}, {ungka}, {ngkai}, {gkait}, {kaitt}, {aitta}, {ittal}, {ttali}, {talib}, {aliba}, {liban}, {ibant}, {bantu}, {antud}, {ntudo}, {tudor}, {udoro}, {doron}, {orong}, {rongt}, {ongto}, {ngtol}, {gtolo}, {tolon}, {olong}, {longt}, {ongto}, {ngtol}, {gtolo}, {tolon}, {olong}, {longm}, {ongmo}, {ngmot}, {gmoti}, {motiv}, {otiva}, {tivas}, {ivasi}, {vasia}, {asias}, {siasu}, {iasaw}, {asawa}, {sawas}, {awasu}, {wasus}, {asusa}, {susah}, {usahc}, {sahce}, {ahcep}, {hcepa}, {cepat}, {epatm}, {patmu}, {atmud}, {tmuda}, {mudah}, {udahu}, {dahup}, {ahupa}, {hupay}, {upaya}, {payad}, {ayada}, {yaday}, {aday}, {daya'}

Setiap kata yang diperoleh dari proses K-gram dilakukan hashing. Sebagai contoh, hasil *hashing* dari kata sesuai dengan hasil *text preprocessing* sebelumnya ditunjukkan pada Tabel VII. Selanjutnya akan dibandingkan dengan setiap data training, salah satu contoh dari data training yaitu sebagai berikut :

” Mohon pihak terkait untuk memperbaiki lampu penerangan jalan yang sudah mati sekitar setahun lebih tidak ditindaklanjuti”

dan mendapatkan hasil teks setelah *preprocessing*, yaitu

” mohon pihak hubung kait kena hubung kait sangkut kait tali meminimalisir pulih optimal tingkat stabil sempurna rekonstruksi lampu penerangan jalan mati tewas berhenti hilang setahun lebih kuat jelas tegas realisasi direspon tuju syah tindak verifikasi lanjut”.

Hasil hashing teks pembanding dapat dilihat pada Tabel VIII.

TABEL VII
 CONTOH HASIL HASHING TEKS

Kata	hash	Kata	hash	Kata	hash	Kata	hash	Kata	hash
lampu	749	rugig	511	itgan	664	nghub	331	olong	718
amput	789	ugiga	976	tgang	252	ghubu	502	longm	494
mpute	643	gigan	145	gangg	800	hubun	515	ongmo	6
puter	357	igang	23	anggu	594	ubung	911	ngmot	552
utera	281	gangg	800	nggum	293	bungk	694	gmoti	23
teran	315	anggu	594	gguma	216	ungka	286	motiv	191
erang	234	nggua	281	gumal	525	ngkai	345	otiva	197
rangj	25	gguan	145	umala	663	gkait	599	tivas	881
angja	595	guanc	19	malam	974	kaits	192	ivasi	170
ngjal	299	uanca	142	alamh	650	aitsa	587	vasia	816
gjala	258	ancam	348	lamha	673	itsan	245	asiaw	353
jalan	821	ncamr	590	amhar	255	tsang	340	siawa	608
alanm	662	camru	301	mhari	937	sangk	413	iawas	978
lanma	757	amrus	886	harim	396	angku	622	awasu	450
anmat	845	mrusa	311	arimo	86	ngkut	496	wasus	298
nmati	32	rusak	40	rimoh	760	gkutk	640	asusa	38
matit	422	usakb	701	imoho	719	kutka	461	susah	424
atite	811	sakbe	225	mohon	644	utkai	464	usaha	679
titew	815	akbeb	294	ohonp	362	tkait	595	sahac	69
itewa	707	kbeba	195	honpi	12	kaitt	193	ahace	212
tewas	565	beban	621	onpih	412	aitta	594	hacep	643
ewasb	972	ebans	800	npiha	354	ittal	292	acepa	794
wasbe	151	bansu	977	pihak	653	ttali	671	cepat	693
asber	33	ansul	264	ihakk	349	talib	707	epatm	603
sberh	389	nsuli	1000	hakka	55	aliba	958	patmu	605
berhe	438	sulit	149	akkai	714	liban	828	atmud	6
erhen	521	ulitb	769	kkait	133	ibant	335	tmuda	193
rhent	30	litba	697	kaitk	184	bantu	984	mudah	388
henti	638	itbah	413	aitke	535	antud	305	udahu	589

Hashing

Hasing merupakan proses mengubah pola (pattern) dan bagian-bagian teks menjadi nilai hash yang merepresentasikan kumpulan karakter tersebut [19]. Fungsi hash yang digunakan mengonversi urutan karakter menjadi bilangan bulat yang unik menggunakan persamaan (1).

$$h(s) = (s[i] \times b^{(n-1)} + s[i + 1] \times b^{(n-2)} + \dots + s[i + n - 1]) \text{ mod } q. \quad (1)$$

dengan

- s = ASCII string yang dicari
- i = nilai iterasi
- b = angka prima
- n = number pada string yang sedang dicari
- q = modulus.

Sebagai contoh, perhitungan untuk tahapan *hasing* adalah sebagai berikut.

$$h_{(lampu)} = ((108 \times 7^{(5-1)}) + (97 \times 7^{(5-2)}) + (109 \times 7^{(5-3)}) + (112 \times 7^{(5-4)}) + (117 \times 7^{(5-5)})) \text{ mod } 1007$$

$$h_{(lampu)} = ((108 \times 7^4) + (97 \times 7^3) + (109 \times 7^2) + (112 \times 7^1) + (117 \times 10^0)) \text{ mod } 1007$$

$$h_{(lampu)} = ((108 \times 2401) + (97 \times 343) + (109 \times 49) + (112 \times 7) + (117 \times 1)) \text{ mod } 1007$$

$$h_{(lampu)} = ((259308) + (33271) + (5341) + (784) + (117)) \text{ mod } 1007$$

$$h_{(lampu)} = 298821 \text{ mod } 1007$$

$$h_{(lampu)} = 749$$



entih	766	tbaha	503	itken	888	ntudo	286	dahup	459
ntihi	727	bahay	561	tken	807	tudor	184	ahupa	272
tihil	244	ahaya	362	kenah	658	udoro	332	hupay	65
ihila	738	hayas	689	enah	848	doron	672	upaya	776
hilan	777	ayasu	129	nahub	287	orong	762	payad	749
ilang	731	yasul	58	ahubu	194	rongt	809	ayada	4
langr	732	asuli	1004	hubun	515	ongto	55	yaday	203
angru	671	sulit	149	ubung	911	ngtol	887	adaya	1004
ngrug	826	ulitg	774	bungh	691	gtolo	360		
grugi	934	litga	732	unghu	285	tolon	528		

TABEL VII
CONTOH HASIL TEKS PEMBANDING

Kata	hash	Kata	hash	Kata	hash	Kata	hash	Kata	hash	Kata	hash
mohon	644	gkutm	642	imalk	57	ruksi	779	arsul	629	sasid	665
ohonp	362	kutme	479	malku	45	uksil	849	rsulu	546	asidi	360
honpi	12	utmern	594	alkur	199	ksila	250	sulut	233	sidir	674
onpih	412	tmemi	487	lkura	537	silam	1005	uluts	367	idire	419
npiha	354	memin	438	kuran	306	ilamp	733	lutse	908	dires	563
pihak	653	emini	927	urang	384	lampu	749	utset	895	iresp	8
ihakh	346	minim	851	rangt	35	ampup	785	tseta	572	respo	703
hakhu	54	inima	789	angti	673	mpupe	615	setah	20	espon	319
akhub	700	nimal	125	ngtin	847	pupen	157	etahu	892	spont	630
khubu	36	imali	55	gting	72	upene	899	tahun	607	pontu	127
hubun	515	malis	29	tingk	523	pener	617	ahunl	269	ontuj	694
ubung	911	alisi	78	ingka	677	enera	87	hunle	24	ntuju	334
bungk	694	lisir	714	ngkat	356	neran	7	unleb	490	tujus	521
ungka	286	isirm	537	gkats	675	erang	234	nlebi	766	ujusy	687
ngkai	345	sirmi	372	katst	725	ranga	16	lebih	513	jusya	123
gkait	599	irmin	328	atsta	290	angan	545	ebihk	135	usyah	806
kaitk	184	rmimi	923	tstab	168	nganj	954	bihku	350	syah	975
aitke	535	minim	851	stabi	214	ganja	815	ihkua	899	yahti	516
itken	888	inima	789	tabil	227	anjat	690	hkuat	903	ahatin	187
tken	807	nimal	125	abils	637	njala	953	kuatj	609	htind	456
kenat	670	imalp	62	bilse	586	jalan	821	uatje	489	tinda	492
enata	912	malpu	80	ilsem	549	alanl	661	atjel	762	indak	470
natal	745	alpul	438	lsemp	463	lanla	750	tjela	450	ndakv	923
atali	367	lpuli	204	sempu	802	anlal	788	jelas	184	dakve	602
talit	725	pulih	990	empur	321	nlala	632	elast	238	akver	283
alite	81	ulihh	691	mpurn	638	lalap	590	laste	48	kveri	126
liten	731	lihhi	159	purna	305	alapn	677	asteg	904	verif	130
itent	663	ihhil	679	urnar	941	lapny	886	stega	437	erifi	572
tenta	239	hhila	351	rnare	898	apnya	722	tegas	788	rifik	378
entan	716	hilan	777	narek	674	pnyal	181	egasr	535	ifika	45
ntang	375	ilang	731	areko	883	nyala	56	gasre	113	fikas	966
tangs	808	lango	729	rekon	303	yalak	581	asrea	800	ikasi	425
angsa	658	angop	645	ekons	545	alako	643	sreal	727	kasil	598
ngsan	742	ngopt	657	konst	198	lakob	625	reali	794	asila	408
gsang	344	gopti	758	onstr	646	akoba	909	ealis	961	silan	1006
sangk	413	optim	292	nstru	1005	kobar	489	alisa	70	ilanj	734
angku	622	ptima	521	struk	175	obars	670	lisan	659	lanju	756
ngkut	496	timal	433	truks	968	barsu	166	isasi	148	anjut	838

Matching

Matching atau string matching adalah proses perbandingan nilai hash dari teks input dengan nilai hash pada teks yang dapat dibandingkan dengan menggunakan persamaan 2.

$$S = \frac{2 \times C}{A+B} \tag{2}$$

dengan

S = Nilai Kesamaan

A + B = Jumlah hash pada string A dan string B

C = Jumlah hash yang sama antara string A dan string B

Hasil matching antara teks yang diuji dengan teks pembanding memiliki hasil seperti berikut.

$$S = \frac{2 \times 168}{195 + 221}$$

$$S = \frac{336}{416}$$

$$S = 0,8076$$

Perhitungan Persentase Similarity

Persentase similarity didapatkan dengan pengalihan hasil matching dengan 100% sehingga menghasilkan nilai persentase sebesar 80,76%.

Penentuan Respons

Setelah mendapatkan nilai persentase similarity, setiap data uji twitter akan dibandingkan dengan masing-masing data latihan twitter. Nilai persentase tertinggi akan dijadikan sebagai acuan untuk menentukan respons. Jika nilai persentase melebihi 70%, data tweet tersebut dianggap valid dijadikan sebagai penentu dan hasil pemetaan jawaban untuk tweet tersebut akan dijadikan hasil respons. Berdasarkan dari pengujian dengan data training sejumlah 485 data tweet yang sudah dilakukan didapatkan hasil kesamaan tweet dengan persentase tertinggi adalah 80,76 % dengan tweet

” Mohon pihak terkait untuk memperbaiki lampu penerangan jalan yang sudah mati sekitar setahun lebih tidak ditindaklanjuti”

Hasil pemetaan respons dari data pelatihan yang sebelumnya telah dilakukan pada tweet tersebut menghasilkan respons sebagai berikut:

” Terima kasih telah mengajukan pengaduan. Pengaduan Anda akan dikirim ke aplikasi Sidumas.”

D. Pengujian

Data respons dari hasil metode Rabin-Karp selanjutnya diujikan untuk penentuan pengaruh Word2Vec dan pengujian K-Gram. Data tweet akan mendapatkan respons jika memiliki nilai persentase kesamaan lebih dari 70%. Jika tidak memenuhi syarat tersebut, data tersebut tidak akan menerima respons secara otomatis. Dalam pengujian kesesuaian, respons akan diperiksa oleh Bidang PIP Kabupaten Badung untuk memastikan kebenaran dan kesesuaiannya. Nilai akurasi di

diperoleh dari perbandingan nilai data uji yang benar dengan total data uji sejumlah 110 data. Nilai akurasi menggunakan persamaan 3

$$S = \frac{\text{Jumlah Data Benar}}{\text{Jumlah Semua Data}} * 100 \%$$

Pengujian implementasi Word2Vec

Pengujian yang menggunakan 110 data uji ini akan melihat pengaruh penerapan metode Word2Vec dalam penentuan respons dengan menggunakan metode Rabin-Karp. Hasil akhirnya adalah tingkat perbedaan nilai kemiripan sebelum dan setelah penggunaan Word2Vec yang ditunjukkan pada Tabel VIII.

TABEL VVIII
HASIL PENGUJIAN PENGARUH WORD2VEC

Pengujian	Kesamaan > 70%	Kesesuaian	Akurasi (%)
Tidak Menggunakan Word2Vec	17	13	11,81
Menggunakan Word2Vec	101	82	74,55

Tabel VIII menunjukkan bahwa Word2Vec sangat memengaruhi hasil kemiripan tweet, tingkat akurasi yang diperoleh adalah 74,55%.

Pengujian K-Gram

Pengujian ini dilakukan untuk mengetahui interpretasi yang berbeda dari setiap nilai k pada K-Gram. Untuk hasil pengujian K-Gram dengan jumlah data uji sebanyak 110 dijelaskan pada Tabel IX.

TABEL VVIII
HASIL PENGUJIAN K-GRAM

K-Gram	Kesamaan > 70%	Kesesuaian	Akurasi Kesesuaian (%)
3	101	82	74,55
4	57	49	44,55
5	48	40	36,36
6	45	37	33,64

Berdasarkan hasil pengujian yang telah dilakukan diperoleh semakin tinggi nilai K-Gram yang digunakan, semakin rendah pula nilai kemiripan yang didapatkan.

IV. DISKUSI

Sejumlah sumber referensi dan data pendukung diperlukan dalam pelaksanaan penelitian ini sebagai pijakan utama untuk pengembangan dan perbandingan dengan penelitian sebelumnya. Oleh karena itu, berbagai rujukan penelitian terdahulu yang relevan berikut ini dianggap penting dan perlu diperhatikan.

Luh Ayu Fernita: Analisis Penentuan Respons Twitter...

Penelitian terdahulu yang pernah dilakukan untuk metode Rabin-Karp adalah penelitian [20]. Penelitian tersebut membahas pengembangan media penilaian otomatis menggunakan algoritma Rabin-Karp dengan pendekatan *Synonym Recognition*. Perhitungan akhir skor siswa yang dilakukan secara otomatis menghasilkan nilai rata-rata sebesar 76,28, sedangkan untuk hasil penilaian yang dilakukan secara manual diperoleh nilai rata-rata sebesar 74,15.

Billhaqi, dkk. [21] telah melakukan penelitian tentang *text similarity*. Dalam penelitian ini, diusulkan solusi penilaian otomatis menggunakan algoritma Rabin-Karp dan *Winnowing* untuk menjawab pertanyaan esai dengan tipe deskripsi bebas dan terbatas. Hasil penelitian menunjukkan tingkat akurasi penggunaan algoritma Rabin-Karp lebih baik jika dibandingkan dengan Algoritma *Winnowing*. Penggunaan Rabin-Karp menghasilkan perbedaan 27,81%, sedangkan penggunaan *Winnowing* menghasilkan perbedaan 44,32% jika dibandingkan dengan penilaian manual. Penelitian ini memberikan kontribusi penting dalam meningkatkan efisiensi dan tujuan pembelajaran online.

Penelitian yang dilakukan oleh Wijaya, dkk [22] merupakan studi yang membandingkan deteksi *plagiarisme* menggunakan algoritma BM25 dan Rabin-Karp pada artikel berbahasa Indonesia. Kesimpulan dari penelitian ini bahwa dari segi performa, terutama waktu eksekusi, algoritma Rabin-Karp memiliki performa yang lebih baik dibandingkan dengan algoritma BM25.

Berdasarkan beberapa penelitian tersebut, dapat disimpulkan bahwa metode Rabin-Karp memberikan hasil yang baik dalam penggunaannya untuk kemiripan teks. Dengan penambahan Word2Vec sebagai representasi kata, kinerja Rabin Karp meningkat.

V. KESIMPULAN

Metode Rabin-Karp efektif dalam menentukan respons berdasarkan data tweet dengan tingkat kemiripan tertinggi. Penggunaan metode Rabin-Karp yang digabungkan dengan metode Word2Vec menghasilkan tingkat akurasi 74,55%. Hasil pengujian K-Gram menunjukkan bahwa semakin rendah nilai K-Gram, maka nilai *similarity* akan semakin tinggi, begitu juga sebaliknya.

Penelitian ini diharapkan dapat memberikan kontribusi dalam konteks pemerintahan yang responsif terhadap isu-isu masyarakat yang dibahas di Twitter. Untuk itu, hasil yang diperoleh ini perlu divalidasi dengan penambahan data latihan dengan tujuan untuk meningkatkan kesesuaian jawaban.

REFERENSI

- [1] K. Q. Fredlina, K. T. Werthi, N. P. Widiari, and K. L. Subagia, "Sosialisasi dan Pelatihan Perlindungan Data Privasi Bagi Siswa di SMKN 3 Denpasar," *JIP - J. Ilm. Ilmu Pendidik.*, vol. 4, no. 2, 2021, doi: 10.54371/jiip.v4i2.212.
- [2] P. A. Permatasari, L. Linawati, and L. Jasa, "Survei Tentang Analisis Sentimen Pada Media Sosial," *Maj. Ilm. Teknol. Elektro*, vol. 20, no. 2, 2021, doi: 10.24843/mite.2021.v20i02.p01.



- [3] R. H. Dwiwina and K. Y. S. Putri, "The Use of the Auto Base Accounts on Twitter as A Media for Sharing Opinions," *Ultim. J. Ilmu Komun.*, 2021, doi: 10.31937/ultimacomm.v13i1.1603.
- [4] A. A. Pailendra and T. Gunarto, "Analisis Pengaruh Peran Masyarakat dan Pemerintah terhadap Pariwisata Curug Klawas Di Lampung Utara," *E-journal F. Econ. Bus. Entrep.*, vol. 1, no. 3, pp. 237–244, Oct. 2022, doi: 10.23960/efebe.v1i3.40.
- [5] Y. Astuti and I. R. Wulandari, "An arrangement of the number of K-grams in the performance of Rabin Karp algorithm in text adjustment," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 26, no. 3, 2022, doi: 10.11591/ijeecs.v26.i3.pp1388-1394.
- [6] F. Husain and O. Uzuner, "Investigating the Effect of Preprocessing Arabic Text on Offensive Language and Hate Speech Detection," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 21, no. 4, 2022, doi: 10.1145/3501398.
- [7] V. A. Flores, L. Jasa, and L. Linawati, "Analisis Sentimen untuk Mengetahui Kelemahan dan Kelebihan Pesaing Bisnis Rumah Makan Berdasarkan Komentar Positif dan Negatif di Instagram," *Maj. Ilm. Teknol. Elektro*, vol. 19, no. 1, 2020, doi: 10.24843/mite.2020.v19i01.p07.
- [8] S. Juniarsih, E. F. Ripanti, and E. E. Pratama, "Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 3, 2020, doi: 10.26418/justin.v8i3.39118.
- [9] S. Styawati, A. Nurkholis, A. A. Aldino, S. Samsugi, E. Suryati, and R. P. Cahyono, "Sentiment Analysis on Online Transportation Reviews Using Word2Vec Text Embedding Model Feature Extraction and Support Vector Machine (SVM) Algorithm," 2022, doi: 10.1109/ISMODE53584.2022.9742906.
- [10] R. Kosasih and A. Alberto, "Sentiment analysis of game product on shopee using the TF-IDF method and naive bayes classifier," *Ilk. J. Ilm.*, vol. 13, no. 2, 2021, doi: 10.33096/ilkom.v13i2.721.101-109.
- [11] B. R. Lidiawaty, M. E. Zulfaqor, O. Diyantara, and D. R. S. Dewi, "Keywords Generator From Paragraph Text Using Text Mining in Bahasa Indonesia," 2022, doi: 10.1109/IRTM54583.2022.9791753.
- [12] I. D. Wahyono *et al.*, "Shared Nearest Neighbour in Text Mining for Classification Material in Online Learning Using Mobile Application," *Int. J. Interact. Mob. Technol.*, vol. 16, no. 4, 2022, doi: 10.3991/ijim.v16i04.28991.
- [13] S. K. Bharti, R. K. Gupta, S. Patel, and M. Shah, "Context-Based Bigram Model for POS Tagging in Hindi: A Heuristic Approach," *Ann. Data Sci.*, 2022, doi: 10.1007/s40745-022-00434-4.
- [14] Y. Qiao, W. Zhang, X. Du, and M. Guizani, "Malware Classification Based on Multilayer Perception and Word2Vec for IoT Security," *ACM Trans. Internet Technol.*, vol. 22, no. 1, 2022, doi: 10.1145/3436751.
- [15] S. Wang, C. Aggarwal, and H. Liu, "Beyond word2vec: Distance-graph tensor factorization for word and document embeddings," 2019, doi: 10.1145/3357384.3358051.
- [16] K. A. Sekarwati, L. Y. Banowosari, I. M. Wiryana, and D. Kerami, "Pengukuran Kemiripan Dokumen dengan Menggunakan Tools Gensim," *Pros. SNST ke-6 Tahun 2015*, 2015.
- [17] I. Z. Simanjuntak, "Analisa Kombinasi Algoritma Stemming Dan Algoritma Soundex Dalam Pencarian Kata Bahasa Indonesia," *Inf. dan Teknol. Ilm.*, vol. 10, no. 1, 2022.
- [18] S. L. B. Ginting, Y. R. Ginting, S. Sutono, and W. A. Sirait, "Aplikasi Deteksi Kemiripan Kata Menggunakan Algoritma Rabin-Karp," *J. Teknol. dan Inf.*, vol. 12, no. 2, 2022, doi: 10.34010/jati.v12i2.6947.
- [19] A. C. D. Sitepu, "Representasi Teknik Information Retrieval Pada Perhitungan Rabin-Karp Menggunakan Stemming Nazief-Adriani," *JUTISAL J. Tek. Inform. Univers.*, vol. 2, no. 1, 2022.
- [20] E. Hamir, D., & Ekohariadi, "MENGUNAKAN ALGORITMA RABIN KARP UNTUK MATA PELAJARAN INFORMATIKA DI MA KANJENG SEPUH Dimas Tifli Irhami Hamir Ekohariadi Abstrak," *J. Inf. Technol. Educ.*, pp. 19–27, 2018.
- [21] T. T. I. Billhaqqi, G. W. Wicaksono, and C. S. K. Aditya, "Comparison analysis of Rabin-Karp and Winnowing algorithms in automated essay answer assessment system," in *AIP Conference Proceedings*, 2022, vol. 2453, doi: 10.1063/5.0095186.
- [22] I. N. S. W. Wijaya, K. A. Seputra, and W. G. S. Parwita, "Comparison of the BM25 and rabinkarp algorithm for plagiarism detection," in *Journal of Physics: Conference Series*, 2021, vol. 1810, no. 1, doi: 10.1088/1742-6596/1810/1/012032.