

## Improvement of Spelling Correction Accuracy in Indonesian Language through the Application of Hamming Distance Method

Mudawil Qulub<sup>1\*</sup>, Rifqi Hammad<sup>2\*</sup>, Pahrul Irfan<sup>3\*</sup>, Yuliana<sup>4\*\*</sup>

\*Ilmu Komputer, Universitas Bumigora

\*\*Institut Shanti Bhuana

[mudawil@universitasbumigora.ac.id](mailto:mudawil@universitasbumigora.ac.id)<sup>1</sup>, [rifqi.hammad@universitasbumigora.ac.id](mailto:rifqi.hammad@universitasbumigora.ac.id)<sup>2</sup>, [irfan@universitasbumigora.ac.id](mailto:irfan@universitasbumigora.ac.id)<sup>3</sup>,  
[yuliana@shantibhuana.ac.id](mailto:yuliana@shantibhuana.ac.id)<sup>4</sup>

### Article Info

#### Article history:

Received 2023-07-19

Revised 2023-11-01

Accepted 2023-11-07

#### Keyword:

*Hamming Distance,  
Spelling Correction,  
Indonesian Language.*

### ABSTRACT

Spelling correction is a critical feature in software to reduce typing errors, commonly found in document processing software and smartphone keyboards. This research aims to evaluate the accuracy of the Hamming Distance method in correcting words in the Indonesian language, both standard and non-standard forms. The research data is derived from a previous study and comprises 60 standard and non-standard Indonesian words. Typos are generated by considering the layout of letters on the QWERTY keyboard. Typing error data is divided into two groups, namely words with 1 and 2 character differences. The first test is conducted on standard words, achieving an accuracy rate of 98.33% for 1 and 2 character differences. Subsequent testing on non-standard words shows an accuracy rate of 100% for 1 character difference and 96.67% for 2 character differences. The results of this research highlight the potential of the Hamming Distance method in improving the quality of spelling correction in the Indonesian language.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

### I. PENDAHULUAN

Tata Bahasa adalah salah satu fitur terpenting dalam Bahasa manusia yang memberikan pemahaman dan minat terhadap teks bagi pembaca ataupun pendengar [1]. Bahasa sebagai alat komunikasi, digunakan untuk menyampaikan buah pikiran, pendapat, atau suatu hal yang dirasakan dan dipikirkan oleh seseorang ke orang lain. Pikiran dan pendapat tersebut akan mudah diterima oleh orang lain jika Bahasa yang digunakan jelas, tepat dan tidak menimbulkan penafsiran ganda. Bahasa Indonesia sebagai Bahasa resmi disemua pendidikan sekolah, mulai dari taman kanak-kanan sampai perguruan tinggi, dimana berfungsi sebagai sarana pengembangan dan pemanfaatan ilmu pengetahuan dan teknologi [2].

Berdasarkan KBBI (Kamus Besar Bahasa Indonesia) Bahasa Indonesia terbagi menjadi 2 yaitu bahasa baku dan tidak baku. Bahasa baku adalah versi bahasa yang diangkat dan disepakati sebagai ragam Bahasa yang dijadikan tolak ukur sebagai Bahasa yang baik dan benar dalam komunikasi bersifat resmi, baik secara lisan maupun tulisan. Sedangkan

Bahasa tidak baku adalah variasi Bahasa yang memakai kata, struktur kalimat, ejaan, pengucapan atau lafal yang tidak benar yang digunakan dalam situasi yang tidak resmi [3].

Bahasa Indonesia sebagai Bahasa nasional digunakan lebih dari 222 juta orang untuk berkomunikasi sekolah, kantor, kehidupan sehari-hari dan sebagainya [4]. Terutama dalam penulisan dokumen, jurnal ilmiah, atau tugas akhir. Namun, dalam penulisan dokumen-dokumen tersebut terdapat kendala yang sering terjadi yaitu kesalahan pada penulisan kata Ketika pengguna sedang menggunakan *keyboard* komputer atau *smartphone* [5]. Beberapa faktor yang menjadi penyebab kesalahan ketik diantaranya yaitu letak huruf *keyboard* yang berdekatan sehingga menyebabkan slip jari tangan [6].

Para peneliti sebelumnya membagi kesalahan ketik menjadi 2 kategori, yaitu deteksi kesalahan dan koreksi [7]. Kesalahan ketik memiliki empat kesaamaan, yaitu penyisipan karakter, permutasi, penggantian dan penghapusan [8]. Dalam mendeteksi kesalahan pengetikan dibutuhkan program komputer yang disebut *Spell checker*. *Spell checker* adalah sebuah proses dalam program komputer untuk memeriksa

setiap ejaan kata dalam teks elektronik atau dokumen dalam urutan yang benar [9].

Penelitian yang relevan terkait kesalahan ketik yaitu penelitian yang berjudul Identifikasi Kesalahan Penulisan Kata (Typographical Error) pada Dokumen Berbahasa Indonesia Menggunakan Metode N-gram dan Levenshtein Distance. Pengujian dilakukan dengan membagi jenis kesalahan penulisan kata menjadi 2 yaitu, *insertion* dan *substitution*. Dimana hasil untuk jenis *insertion* didapatkan nilai presisi sebesar 0.97. Sedangkan untuk jenis *substitution* mendapat nilai *recall* sebesar 1 [4].

Penelitian selanjutnya terkait kesalahan ketik pada Bahasa Indonesia menggunakan metode *Burkhard Keller Tree* dan *SymSpell*. Dimana hasil dari pengujian kedua metode memiliki tingkat *accuracy*, *precision*, dan *recall* sama yaitu 1, dikarenakan kedua metode menghasilkan hasil koreksi kesalahan kata yang sama [10].

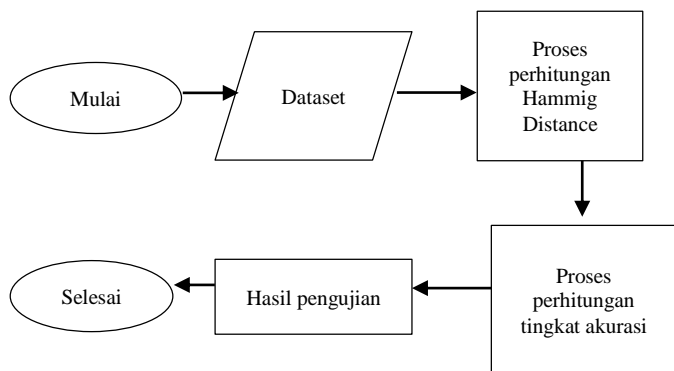
Salah satu algoritma yang digunakan untuk pencocokan kata adalah algoritma Hamming Distance. Algoritma Hamming Distance adalah salah satu algoritma pencocokan string yang digunakan dalam teori informasi sebagai estimasi kesalahan kata dengan cara membandingkan simbol-simbol pada dua string yang memiliki Panjang yang sama [11].

Berdasarkan paparan diatas, dalam penelitian ini dilakukan penerapan dan analisa metode Hamming Distance terhadap kesalahan kata baku dan tidak baku dalam Bahasa Indonesia, untuk mengetahui tingkat akurasi lebih baik pada koreksi kesalahan kata.

## II. METODE PENELITIAN

### A. Alur Penelitian

Alur penelitian berikut memberikan gambaran penelitian meliputi langkah-langkah yang ditempuh, sumber data, prosedur, yang kemudian diolah dan dianalisis.

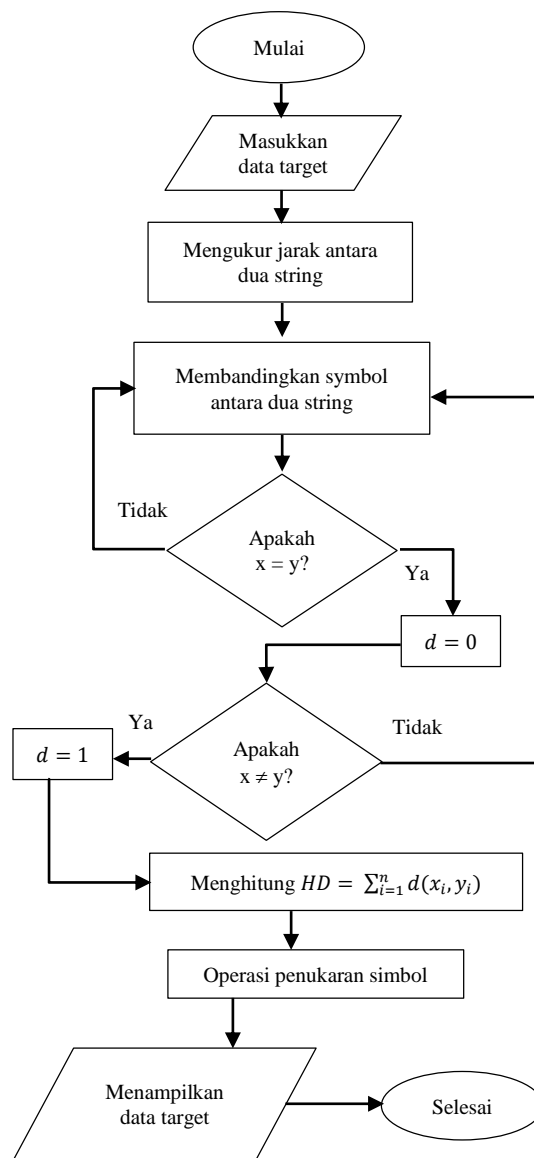


Gambar 1. Alur penelitian

### B. Algoritma Hamming Distance

Hamming Distance adalah algoritma pengecekan kata dengan cara membandingkan dua *string* yang panjangnya sama, kemudian menghitung jarak kesalahan minimal dengan melakukan substitusi karakter yang berbeda sehingga

memperoleh *string* yang tepat. Berikut dibawah ini adalah flowchart dari algoritma *Hamming Distance* [14].



Gambar 2. Flowchart Hamming Distance

### C. Pengumpulan Data

Data yang penulis pada penelitian ini merupakan data kata baku dan kata tidak baku bahasa Indonesia yang berdasar penelitian “Non-formal Affixed Word Stemming in Indonesian Language” [12]. Untuk jumlah data kata yang digunakan pada data kata baku dan tidak baku masing-masing berjumlah 60 kata.

Data kata kesalahan ketik yang penulis gunakan dalam penelitian ini akan disimulasikan dengan memperhatikan tata letak huruf pada layout keyboard QWERTY. Berikut dibawah ini adalah tabel 1 menampilkan tata letak keyboard QWERTY [13] sebagai acuan untuk membentuk data kesalahan kata.

TABEL 1  
TATA LETAK LAYOUT KEYBOARD QWERTY

Huruf	Huruf Tetangga	Jumlah Huruf Tetangga
A	SQZWX	5
B	NVGHF	5
C	VXDFS	5
D	FSECRVXW	8
E	RWDFS	5
F	GDRVTBCE	8
G	HFTBYNVR	8
H	JGYNMBT	8
I	OUKLJ	5
J	KHUMINY	7
K	LJIOMU	6
L	KOPI	4
M	NJKH	4
N	MBHJG	5
O	PILK	4
P	OL	2
Q	WAS	3
R	TEFGD	5
S	DAWXECZQ	8
T	YRGHF	5
U	IYJKH	5
V	BCFGD	5
W	EQSDA	5
X	CZSDA	5
Y	UTHJG	5
Z	XAS	3

Pada data kata kesalahan ketik dibagi menjadi dua bagian yaitu nilai perbedaan karakter 1 dan 2 huruf.

D. Metode Analisis Data

Metode analisis data pada penelitian ini menggunakan metode *Confusion Matrix*. *Confusion Matrix* pada dasarnya memberikan perbandingan hasil klarifikasi yang dilakukan sistem dengan hasil klarifikasi sebenarnya. *Confusion Matrix* berbentuk tabel matrix yang menggambarkan kinerja model klasifikasi pada serangkaian data uji yang nilai sebenarnya diketahui [15]. Tabel 2 dibawah ini menunjukkan tabel dari *confusion matrix*.

TABEL 2  
CONFUSION MATRIX

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Untuk perhitungan *accuracy*, *precision*, dan *error rate* menggunakan rumus pada tabel 4 dibawah ini.

TABEL 3  
RUMUS PRECISSION, ACCURACY, DAN ERROR RATE

Metriks	Rumus
accuracy	$\frac{TP + TN}{TP + FN + TN + FN}$
error rate	$\frac{TP + FN}{TP + FN + TN + FN}$
precision	$\frac{TP}{TP + FN}$

III. HASIL DAN PEMBAHASAN

Berikut merupakan tahapan pengujian yang dilakukan pada algoritma *Hamming Distance* dengan menggunakan data kata baku dan tidak baku bahasa Indonesia. Pada pengujian ini dilakukan pemisahan data kata baku dan tidak baku pada basis data, sehingga setiap basis data memiliki dataset masing-masing 60 kata.

A. Pengujian Kata Baku

Pengujian algoritma *Hamming Distance* dilakukan dengan perbedaan kesalahan kata 1 dan 2 karakter. Hasil pengujian disajikan dalam tabel untuk mengetahui keberhasilan pengecekan tiap katanya.

1) Perbedaan 1 Karakter

Hasil pengujian algoritma *Hamming Distance* pada kata baku bahasa Indonesia dengan nilai *distance* 1 karakter, ditunjukkan pada tabel 4 dibawah ini.

TABEL 4  
HASIL PENGUJIAN KATA BAKU DENGAN PERBEDAAN 1 KARAKTER

No	Kata Baku	Perbedaan 1 karakter	Hasil Ditemukan (Y) dan Tidak Ditemukan (T)
1	terjang	terjsng	Y
2	tuduh	tudyh	Y
3	terima	tetima	Y
4	tegur	tefur	Y
5	pukul	pykul	Y
6	pimpin	pimoin	Y
7	coba	cova	Y
8	siram	sirwm	Y
9	suruh	surih	Y
10	simpan	sinpan	Y
11	sebrang	sebtang	Y
12	anggap	anggps	Y
13	amuk	amhk	Y
14	ambil	ambol	Y
15	buka	buja	Y
16	bantu	bahtu	Y
17	lepas	leoas	Y
18	bayang	bauang	Y
19	injak	injqk	Y
20	kabul	kzbul	Y
21	pergok	perfok	Y
22	tipu	perfok	Y
23	ulang	ukang	Y
24	wujud	wujhd	Y
25	cerita	cetita	Y
26	betul	berul	Y
27	manja	manha	Y
28	ganggu	gwnngu	Y
29	ganti	gajti	Y
30	ikut	ikyt	Y
31	musuh	muduh	T (tuduh)
32	sabun	sagun	Y
33	teman	tejan	Y
34	tukar	tukzr	Y
35	tanya	tanta	Y
36	tunjuk	tubjuk	Y
37	penting	penging	Y
38	pegang	pegahg	Y
39	selamat	selzmat	Y
40	sempat	sempxt	Y

41	korban	korgan	Y
42	hadap	hzdap	Y
43	Bukti	bjkti	Y
44	warna	qarna	Y
45	dengar	denbar	Y
46	ketemu	ketehu	Y
47	benar	brnar	Y
48	begini	beyini	Y
49	kawin	kaqin	Y
50	main	maij	Y
51	parkir	parkur	Y
52	dulu	dilu	Y
53	gendut	grndut	Y
54	karat	karwt	Y
55	paling	paking	Y
56	sabar	sagar	Y
57	bagus	bafus	Y
58	sana	saba	Y
59	cepat	crpat	Y
60	pagi	pqgi	Y

Dari hasil pengujian diatas menunjukkan bahwa algoritma *Hamming Distance* menghasilkan perbaikan kata baku dengan nilai perbedaan 1 karakter yaitu sejumlah 59 kata berhasil dan 1 kata gagal di perbaiki.

## 2) Perbedaan 2 Karakter:

Pengujian berikutnya dilakukan dengan kata salah dengan perbedaan 2 karakter. Dimana hasil pengujian ditunjukkan pada tabel 5.

TABEL 5  
HASIL PENGUJIAN KATA BAKU DENGAN PERBEDAAN 2 KARAKTER

No	Kata Baku	Perbedaan 2 karakter	Hasil Ditemukan (Y) dan Tidak Ditemukan (T)
1	terjang	twrjsng	Y
2	tuduh	tidoh	Y
3	terima	rekma	Y
4	tegur	yegjr	Y
5	pukul	pilyl	Y
6	pimpin	pomoin	Y
7	coba	voga	Y
8	siram	sursm	Y
9	suruh	durjh	Y
10	simpan	sumpsn	Y
11	sebrang	srbrang	Y
12	anggap	snggao	Y
13	amuk	smjk	Y
14	ambil	anbul	Y
15	buka	biks	Y
16	bantu	bamti	Y
17	lepas	lrpqs	Y
18	bayang	bsyang	Y
19	injak	onjsk	Y
20	kabul	kzbyl	Y
21	pergok	pwrgek	Y
22	tipu	topy	Y
23	ulang	ilsng	Y
24	wujud	wijuc	Y
25	cerita	cerots	Y
26	betul	bwjtl	Y
27	manja	mahjs	Y
28	ganggu	gamggi	Y
29	ganti	fsnti	Y
30	ikut	okit	Y

31	musuh	miduh	T (tuduh)
32	sabun	sqbin	Y
33	teman	yemsn	Y
34	tukar	rukwr	Y
35	tanya	tsnyw	Y
36	tunjuk	tinjok	Y
37	penting	pwntkng	Y
38	pegang	oehang	Y
39	selamat	selsmwt	Y
40	sempat	srmoat	Y
41	korban	kirvan	Y
42	Hadap	hsdao	Y
43	bukti	bikri	Y
44	warna	watns	Y
45	dengar	dwngat	Y
46	ketemu	krtenu	Y
47	benar	bwnat	Y
48	begini	vegoni	Y
49	kawin	ksqin	Y
50	main	mson	Y
51	parkir	pskrur	Y
52	dulu	dyli	Y
53	gendut	grmdut	Y
54	karat	ketat	Y
55	paling	pwlung	Y
56	sabar	savwr	Y
57	bagus	bsgis	Y
58	sana	sqma	Y
59	cepat	xcoat	Y
60	pagi	psgo	Y

Dari pengujian kata baku dengan nilai perbedaan 2 karakter menghasilkan 59 kata benar dan 1 kata salah. Dimana kata yang gagal di perbaiki adalah kata “musuh” dari kata yang salah “miduh” menghasilkan kata “tuduh”.

## B. Pengujian Kata Tidak Baku

Pengujian pada tahap ini menggunakan kata bahasa Indonesia tidak baku. Berikut dibawah ini hasil pengujian dengan perbedaan nilai 1 dan 2 karakter.

### 1) Perbedaan 1 Karakter:

Berikut tabel 6 dibawah ini menampilkan hasil pengujian kata tidak baku menggunakan algoritma *Hamming Distance* dengan nilai perbedaaan 1 karakter.

TABEL 6  
HASIL PENGUJIAN KATA TIDAK BAKU DENGAN PERBEDAAN 1 KARAKTER

No	Kata Tidak Baku	Perbedaan 1 karakter	Hasil Ditemukan (Y) dan Tidak Ditemukan (T)
1	nerjang	netjang	Y
2	nuduh	njduh	Y
3	nerima	neroma	Y
4	negur	negyr	Y
5	mukul	muluk	Y
6	mimpin	mimpun	Y
7	nyoba	nykba	Y
8	nyiram	nyuram	Y
9	nyuruh	nhuruh	Y
10	nyimpen	nykmpen	Y
11	nyebrang	nywbrang	Y
12	nganggep	nganggwp	Y
13	ngamuk	ngamul	Y
14	ngambil	ngambkl	Y

15	ngebuka	ngwbuka	Y
16	ngebantu	ngeantu	Y
17	ngelepas	ngrelepas	Y
18	kebayang	kebsyang	Y
19	keinjek	keinhek	Y
20	kekabul	kelabul	Y
21	kepergok	kepetgok	Y
22	ketipu	ketiou	Y
23	keulang	keolang	Y
24	kewujud	kewijud	Y
25	critain	crirain	Y
26	betulin	berulin	Y
27	manjain	manjsin	Y
28	gangguin	ganhguin	Y
29	gantian	gantuan	Y
30	ikutan	ikytan	Y
31	musuhan	muduhan	Y
32	sabunan	sanunan	Y
33	temenan	temwnan	Y
34	tukeran	tuleran	Y
35	nanyain	nanyaon	Y
36	nunjukin	nunjukon	Y
37	mentingin	mwntingin	Y
38	megangin	megangkkn	Y
39	nyelametin	nyekametin	Y
40	nyempetin	nydmpetin	Y
41	ngorbanin	ngorbamin	Y
42	Ngadepin	ngafepin	Y
43	ngebuktiin	ngebiktiin	Y
44	ngewarnain	ngewaenain	Y
45	kedengeran	kedenferan	Y
46	ketemuan	keyemuan	Y
47	beneran	bemeran	Y
48	ginian	gimian	Y
49	kawinan	kaqinan	Y
50	mainan	msinan	Y
51	parkiran	parkoran	Y
52	duluan	dulian	Y
53	gendutan	genditan	Y
54	karatan	kawatan	Y
55	palingan	pqlingan	Y
56	sabaran	sabatan	Y
57	kebagusan	kebagudan	Y
58	sanaan	sanaam	Y
59	cepatan	ceprtan	Y
60	sepagian	sepggian	Y

Dari pengujian kata tidak baku algoritma *Hamming Distance* diatas menghasilkan 60 kata berhasil diperbaiki tanpa ada satu kata yang salah.

## 2) Perbedaan 2 Karakter

Pada pengujian data kata tidak baku dengan perbedaan 2 karakter algoritma *Hamming Distance* ditunjukkan pada tabel 7 berikut.

TABEL 7  
HASIL PENGUJIAN KATA TIDAK BAKU DENGAN PERBEDAAN 2 KARAKTER

No	Kata Tidak Baku	Perbedaan 2 karakter	Hasil Ditemukan (Y) dan Tidak Ditemukan (T)
1	nerjang	nsrjsng	Y
2	nuduh	nufih	Y
3	nerima	neroms	Y
4	negur	negie	Y
5	mukul	mjkyl	Y
6	mimpin	mompib	Y
7	nyoba	nypbs	Y
8	nyiram	nyitsm	Y
9	nyuruh	nyiruj	T (nyiram)
10	nyimpen	ntimpwn	Y
11	nyebrang	nywbrsng	Y
12	nganggep	nganffep	Y
13	ngamuk	ngamil	Y
14	ngambil	nganbol	Y
15	ngebuka	ngevika	Y
16	ngebantu	ngwbanti	Y
17	ngelepas	ngelrpd	Y
18	kebayang	kecaysng	Y
19	keinjek	keinjwl	Y
20	kekabul	kekanuk	Y
21	kepergok	kdperfok	Y
22	ketipu	kstipi	Y
23	keulang	kejzng	Y
24	kewujud	kewijux	Y
25	critain	criyaon	Y
26	betulin	bwtilin	Y
27	manjain	mznjsin	Y
28	gangguin	ganfgoin	Y
29	gantian	gznrian	Y
30	ikutan	okytan	Y
31	musuhan	mudujan	Y
32	sabunan	saguman	Y
33	temenan	trmenzn	Y
34	tukeran	tiketn	Y
35	nanyain	nantaon	Y
36	nunjukin	ninjykin	Y
37	mentingin	mwntingkn	Y
38	megangin	mefamgin	Y
39	nyelametin	nyekamptin	Y
40	nyempetin	nywmptrin	Y
41	ngorbanin	ngorvsnim	Y
42	ngadepin	ngsdwpin	Y
43	ngebuktiin	nfebukyiin	Y
44	ngewarnain	ngeqarmain	Y
45	kedengeran	kefengerzn	Y
46	ketemuan	krtemian	Y
47	beneran	banetan	Y
48	ginian	gonisn	Y
49	kawinan	kswiman	Y
50	mainan	msonan	Y
51	parkiran	psrliran	Y
52	duluan	diluzn	Y
53	gendutan	gensuran	Y
54	karatan	kstatan	Y
55	palingan	pwlungan	Y
56	sabaran	sanatan	T (karatan)
57	kebagusan	krbafusan	Y
58	sanaan	sanssn	Y
59	cepatan	crpwtan	Y
60	sepagian	srpahian	Y

Hasil pengujian kata tidak baku dengan algoritma *Hamming Distance* dengan perbedaan 2 karakter menghasilkan 58 kata berhasil diperbaiki sesuai target dan 2 kata gagal diperbaiki.

### C. Perhitungan Persentase Keberhasilan

Hasil dari setiap pengujian memiliki persentase keberhasilan yang hampir sama dalam melakukan perbaikan kata terhadap kata baku dan tidak baku dengan perbedaan 1 dan 2 karakter. Perbandingan tingkat akurasi dihitung berdasar pada jumlah kata benar dan kata salah pada tabel 4, 5, 6, dan 7.

#### 1) Persentase Keberhasilan Kata Baku

Hasil persentase tingkat keberhasilan koreksi kesalahan kata baku dengan perbedaan 1 dan 2 karakter ditampilkan pada tabel 8 berikut ini.

TABEL 8  
PERBANDINGAN TINGKAT KEBERHASILAN PENGUJIAN KATA BAKU

Nama Algoritma	Benar	Salah	Akurasi
Hamming Distance dengan perbedaan 1 karakter	59	1	98,33%
Hamming Distance dengan perbedaan 2 karakter	59	1	98,33%

Dari persentase keberhasilan algoritma *Hamming Distance* untuk koreksi kesalahan kata baku dengan perbedaan 1 dan 2 karakter menghasilkan tingkat persentase yang sama yaitu 98,33%. Kata yang gagal diperbaiki adalah kata "musuh". Dimana kata salah yang digunakan dengan perbedaan 1 karakter adalah "mudah" dan perbedaan 2 karakter yaitu "miduh". Dari kedua kesalahan kata tersebut menghasilkan kata "tuduh".

#### 2) Persentase Keberhasilan Kata Tidak Baku

Hasil persentase pengujian pada kata tidak baku dengan perbedaan 1 dan 2 karakter ditampilkan pada tabel 9 berikut.

TABEL 9  
PERBANDINGAN TINGKAT KEBERHASILAN PENGUJIAN KATA TIDAK BAKU

Nama Algoritma	Benar	Salah	Akurasi
Hamming distance dengan perbedaan 1 karakter	60	0	100%
Hamming distance dengan perbedaan 2 karakter	58	2	96,67%

Dari hasil presentasi pengujian algoritma pada kata tidak baku menghasilkan tingkat akurasi 100% dengan perbedaan 1 karakter. Sedangkan untuk perbedaan 2 karakter menghasilkan tingkat akurasi 96,67%, dimana 58 kata berhasil dan 2 kata gagal.

### D. Pengukuran dengan Confusion Matrix

#### 1) Kata Baku Dengan Perbedaan 1 Karakter

Hasil *confusion matrix* ditampilkan pada tabel 10 berikut.

TABEL 10  
CONFUSION MATRIX KATA BAKU DENGAN NILAI DISTANCE 1

Kata Baku	Diproses Benar	Tidak Diproses
Berhasil	59	1
Gagal	0	0

Berdasarkan data pada tabel, pengujian algoritma *Hamming Distance* kata baku dengan perbedaan 1 karakter menghasilkan tingkat *accuracy* 0,983, *precision* 1, dan *error rate* 0,016.

#### 2) Kata Baku Perbedaan 2 Karakter

Hasil *confusion matrix* kata baku dengan perbedaan 2 karakter ditunjukkan pada tabel 11 dibawah ini.

TABEL 11  
CONFUSION MATRIX KATA BAKU DENGAN NILAI DISTANCE 2

Kata Baku	Diproses Benar	Tidak Diproses
Berhasil	59	1
Gagal	0	0

Berdasarkan data pada tabel 13, hasil pengujian kata baku dengan kata salah dengan perbedaan 2 karakter menghasilkan hasil yang sama dengan 1 karakter. Dimana menghasilkan tingkat yaitu *accuracy* 0,983, *precision* 1, dan *error rate* 0,016.

#### 3) Kata Tidak Baku dengan Perbedaan 1 Karakter

Hasil *confusion matrix* untuk kata tidak baku dengan perbedaan 1 karakter ditampilkan pada tabel 12 berikut.

TABEL 12  
CONFUSION MATRIX KATA TIDAK BAKU DENGAN NILAI DISTANCE 1

Kata Baku	Diproses Benar	Tidak Diproses
Di-Autocorrect	60	0
Tidak Di-Autocorrect	0	0

Berdasarkan data tabel 12, hasil pengujian kata tidak baku dengan perbedaan 1 karakter, menghasilkan tingkat *accuracy* 1, *precision* 1, dan *error rate* 0.

#### 4) Kata Tidak Baku dengan Perbedaan 2 Karakter

Hasil *confusion matrix* untuk kata tidak baku dengan perbedaan 2 karakter ditampilkan pada tabel 13 berikut.

TABEL 13  
CONFUSION MATRIX KATA TIDAK BAKU DENGAN NILAI DISTANCE 2

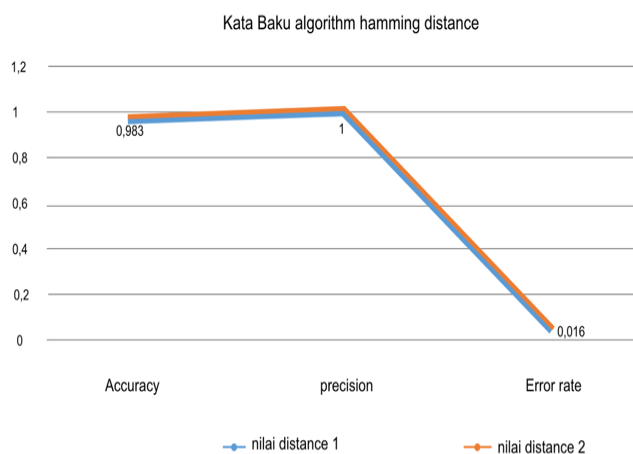
Kata Baku	Diproses Benar	Tidak Diproses
Di-Autocorrect	58	2
Tidak Di-Autocorrect	0	0

Berdasarkan data tabel 13, hasil pengujian kata tidak baku dengan perbedaan 2 karakter, menghasilkan tingkat *accuracy* 0,966, *precision* 1, dan *error rate* 0,033.

### E. Grafik Hasil Perhitungan Confusion Matrix

#### 1) Grafik Kata Baku

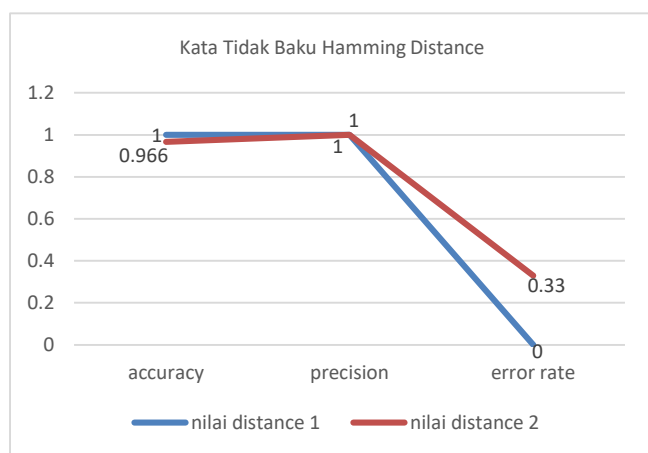
Grafik hasil pengujian algoritma *Hamming Distance* dengan dengan perbedaan karakter 1 dan 2 ditampilkan pada gambar 3 berikut.



Gambar 3. Kata Baku Algoritma *Hamming Distance*

#### 2) Grafik Kata Tidak Baku

Grafik hasil pengujian kata tidak baku algoritma *Hamming Distance* dengan dengan perbedaan 1 dan 2 karakter ditampilkan pada gambar 4 berikut.



Gambar 4. Kata Tidak Baku Algoritma *Hamming Distance*

### IV. KESIMPULAN

Berdasarkan hasil analisa dan evaluasi yang dilakukan terhadap pengujian, dimana pengujian dilakukan dengan memisahkan kata baku dan tidak baku pada basis data dapat disimpulkan sebagai berikut. Pada pengujian kata baku, skenario dilakukan dengan kata salah memiliki perbedaan 1 dan 2 karakter, dimana algoritma *Hamming Distance* menghasilkan tingkat akurasi yang sama yaitu 98,33%.

Pengujian berikutnya pada kata tidak baku, skenario pertama pengujian menggunakan kata salah dengan perbedaan 1 karakter menghasilkan tingkat akurasi 100%. Kemudian skenario kedua, pengujian kata salah dengan perbedaan 2 karakter menghasilkan tingkat akurasi sebesar 96,67%.

### DAFTAR PUSTAKA

- [1] A. Musyafa, Y. Gao, A. Solyman, C. Wu, and S. Khan, "Automatic Correction of Indonesian Grammatical Errors Based on Transformer," *Appl. Sci.*, vol. 12, no. 20, p. 10380, Oct. 2022, doi: 10.3390/app122010380.
- [2] A. Muchti and Y. Ernawati, "Penguasaan Kosakata Baku Dan Tidak Baku: Sebuah Studi Kasus Mahasiswa UBD," *J. Ilm. Bina Edukasi*, vol. 15, no. 1, pp. 61–70, Jun. 2022, doi: 10.33557/jedukasi.v15i1.1762.
- [3] D. Ginting, "Kemampuan Membedakan Bahasa Indonesia Baku dan Tidak Baku Oleh Siswa (Studi Kasus Siswa SMP Negeri 3 Mardinding)," no. 1.
- [4] A. I. Fahma, I. Cholissodin, and R. S. Perdana, "Identifikasi Kesalahan Penulisan Kata (Typographical Error) pada Dokumen Berbahasa Indonesia Menggunakan Metode N-gram dan Levenshtein Distance".
- [5] Cahyo Prianto, D. Markuci, and S. F. Pane, "Implementasi Spelling Corrector Untuk Mengatasi Typographical Error Pada Fitur Pencarian Aplikasi Kamus Istilah Informatika," *J. Teknol. Inf. J. Keilmuan Dan Apl. Bid. Tek. Inform.*, vol. 17, no. 1, pp. 20–30, Jan. 2023, doi: 10.47111/jti.v17i1.5520.
- [6] S. N. Agustin, I. Mufarrihah, and D. R. Prehanto, "Aplikasi Pengoreksian Kesalahan Bahasa Indonesia Menggunakan Metode Jaro Winkler Distance Dan Levenshtein Distance Berbasis Web".
- [7] W. L. Wong, M. M. Muhammad, K. P. Chuah, N. Saimi, A. H. Ma'arop, and R. Elias, "Did you Run the Telegram? Use of Mobile Spelling Checker on Academic Writing," *Multiling. Acad. J. Educ. Soc. Sci.*, vol. 10, no. 1, p. Pages 1-19, Jan. 2022, doi: 10.46886/MAJESS/v10-i1/7379.
- [8] K. Goslin and M. Hofmann, "English Language Spelling Correction as an Information Retrieval Task Using Wikipedia Search Statistics".
- [9] D. A. Anggoro and I. Nurfadilah, "Active Verb Spell Checking Mem + P in Indonesian Language Using the Jaro-Winkler Distance Algorithm," *Iraqi J. Sci.*, pp. 1811–1822, Apr. 2022, doi: 10.24996/ijss.2022.63.4.38.
- [10] M. H. Ferdiansyah and I. K. D. Nuryana, "Analisis Perbandingan Metode Burkhard Keller Tree dan SymSpell dalam Spell Correction Bahasa Indonesia," vol. 04, 2023.
- [11] Y. N. Gulo, "Penerapan Algoritma Hamming Distance Untuk Pencarian Teks Pada Aplikasi Ensiklopedia Indonesia," vol. 1, no. 2, 2022.
- [12] R. B. S. Putra and E. Utami, "Non-formal affixed word stemming in Indonesian language," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, Yogyakarta: IEEE, Mar. 2018, pp. 531–536. doi: 10.1109/ICOIACT.2018.8350735.
- [13] A. Kirk, "Improving the Accuracy of Mobile Touchscreen QWERTY Keyboards".
- [14] T. K. Wulandari, E. D. Oktaviani, and A. Lestari, "Penerapan Metode Binary Search dan Hamming Distance pada E-library SMAN 2 Katingan Hilir," vol. 2, 2022.
- [15] D. S. Suparno, "Pengenalan Pola Untuk Mengetahui Jumlah Target Pengunjung Mall Berdasarkan Usia, Gender, Pendapatan Tahun, Pengeluaran, Tujuannya Untuk Mempermudah Mengetahui Target Pasar Menggunakan Metode EDA, K-Means, Hierarchical Clustering, Confusion Matrix," vol. 3, no. 2, 2021.