



## Optimasi Metode *Multinomial Naïve Bayes* Dengan Menggunakan Metode *Levenshtein Distance* Pada Ulasan Aplikasi KAI Access

Muhammad Ivan Fariz<sup>1\*</sup>, Deni Arifianto<sup>2</sup>, Yeni Dwi Rahayu<sup>3</sup>

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember<sup>1,3</sup>

Sistem Informasi, Fakultas Teknik, Universitas Muhammadiyah Jember<sup>2</sup>

Email: mivanfariz2001@gmail.com<sup>1\*</sup>, deniarifianto@unmuhjember.ac.id<sup>2</sup>, yenedwirahayu@unmuhjember.ac.id<sup>3</sup>

### ABSTRAK

*KAI Access* merupakan aplikasi untuk mempermudah pengguna dalam mengakses layanan dan informasi terkait tiket perjalanan kereta api. *KAI Access* memiliki fitur ulasan beraneka ragam yang merupakan wadah bagi pengguna untuk memberikan *feedback*. Pada penelitian ini akan dilakukan analisis sentimen terhadap data ulasan *KAI Access* sejumlah 8713 data ulasan. Metode yang digunakan yaitu *Multinomial Naïve Bayes* dan metode *Levenshtein Distance*. Dari kombinasi metode tersebut diharapkan dapat meningkatkan hasil klasifikasi. Seluruh data akan dilakukan pemodelan menggunakan metode *K Fold Cross Validation* dengan nilai  $k=2, 3, 4, 5, 6, 7, 8, 9$  dan  $10$ . Berdasarkan hasil pengujian menggunakan *Multinomial Naïve Bayes* diperoleh nilai tertinggi yaitu akurasi sebesar 92%, tingkat presisi sebesar 60% dan tingkat *recall* sebesar 45%. Sedangkan kombinasi metode *Multinomial Naïve Bayes-levenshtein distance* mengalami peningkatan pada pengujian  $K$ -ke 9 yang awalnya menggunakan *Multinomial Naïve Bayes* saja mendapatkan nilai akurasi 83%, presisi 59% dan *recall* 44%, kemudian meningkat ketika menggunakan kombinasi metode *Levenshtein Distance* sebesar akurasi 84%, presisi 60% dan *recall* 46%. Proses pengujian yang telah dilakukan dapat disimpulkan bahwa optimasi *Multinomial Naïve Bayes* menggunakan *Levenshtein Distance* dapat meningkatkan hasil akurasi, presisi dan *recall*.

**Kata Kunci:** *KAI Access*, Analisis Sentimen, *Multinomial Naïve Bayes*, *Levenshtein Distance*

### ABSTRACT

*KAI Access* is an application to facilitate users in accessing services and information related to train travel tickets. *KAI Access* has a diverse review feature which is a forum for users to provide *feedback*. In this research, sentiment analysis will be carried out on *KAI Access* review data totaling 8713 review data. The methods used are *Multinomial Naïve Bayes* and *Levenshtein Distance* method. The combination of these methods is expected to improve classification results. All data will be modeled using the *K Fold Cross Validation* method with a value of  $k = 2, 3, 4, 5, 6, 7, 8, 9$  and  $10$ . Based on the test results using *Multinomial Naïve Bayes*, the highest value is obtained, namely accuracy of 92%, precision level of 60% and recall level of 45%. While the combination of the *Multinomial Naïve Bayes-levenshtein distance* method experienced an increase in the  $K$ -9th test which initially used *Multinomial Naïve Bayes* alone to get an accuracy value of 83%, precision of 59% and recall of 44%, then increased when using a combination of *Levenshtein Distance* methods of 84% accuracy, 60% precision and 46% recall. The testing process that has been carried out can be concluded that *Multinomial Naïve Bayes* optimization using *Levenshtein Distance* can improve accuracy, precision and recall results.

**Keywords:** *KAI Access*, Sentiment Analysis, *Multinomial Naïve Bayes*, *Levenshtein Distance*

## 1. PENDAHULUAN

Perkembangan teknologi informasi pada zaman sekarang sudah mengalami peningkatan yang memberikan perkembangan modern untuk masyarakat. PT. Kereta Api juga turut memanfaatkan perkembangan teknologi informasi salah satunya yaitu terdapat aplikasi pemesanan tiket kereta api secara online yaitu *KAI Access* yang tersedia pada layanan *Google Play* (Suwondo dkk., 2017). Situs *Google Play* terdapat fitur ulasan yang merupakan wadah bagi pengguna untuk memberikan *feedback* pada aplikasi *KAI Access*, jumlah ulasan aplikasi *KAI Access* saat ini sudah mencapai 167.000 ulasan. Aplikasi pada *google play* terdapat ulasan yang ambigu, yang ditandai dengan ketidaksamaan dari rating dan komentar yang tidak sesuai. Sehingga pada ulasan aplikasi tersebut diperlukan analisis klasifikasi teks untuk mendeteksi kesesuaian rating dan komentar.

Salah satu penerapan metode yang dapat digunakan untuk melakukan klasifikasi suatu data adalah metode *Naïve Bayes* (Lestari & Saepudin, 2021). Turunan dan model dari algoritma *naïve bayes* adalah *Multinomial Naïve Bayes*, metode *Multinomial Naïve Bayes* dipilih karena adapat melakukan

komputasi dengan cepat pada klasifikasi data dan dapat menyelesaikan masalah yang memiliki *multiple class*. Penelitian oleh Yuyun dan Nurul Hidayah yang mengklasifikasi sentimen masyarakat terhadap penanganan covid-19 pada *twitter* menggunakan metode *multinomial naïve* dengan pembobotan TF-IDF memperoleh nilai akurasi, presisi dan recall sebesar 74%, peneliti menyebutkan perlu ditambahkan algoritma lain untuk membantu kinerja algoritma *Multinomial Naïve Bayes* dan meningkatkan nilai akurasi, presisi dan recall pada proses klasifikasi sentimen (Yuyun dkk., 2021).

Kemudian penelitian lainnya oleh Machrus Izunnahadi, Ginanjar Abdurrahman dan Ari Eko Wardoyo yang mengklasifikasi ulasan aplikasi *KAI Access* menggunakan metode *Multinomial Naïve Bayes* dengan 7500 data memperoleh nilai akurasi yaitu sebesar 89%, presisi 92% dan recall 85%, pada penelitian ini terdapat suatu ulasan dari aplikasi yang tidak baku dan kesalahan ejaan pada kata sehingga perlu ditambahkan metode tambahan untuk menemukan hasil yang lebih baik (Izunnahdi dkk., 2023). Ulasan aplikasi pada *Google Play* terdapat suatu bahasa dalam bentuk yang tidak baku, singkatan suatu kata dan kesalahan ejaan kata. Permasalahan kata yang tidak baku pada ulasan *Google Play* dapat dilakukan perbaikan, perbaikan kata tidak baku tersebut dapat dilakukan normalisasi kata menggunakan metode *Levenshtein Distance*.

Dari studi kasus yang terdapat pada penelitian terdahulu, klasifikasi suatu data dengan metode *Multinomial Naïve Bayes* belum optimal dikarenakan terdapat ulasan dengan kata yang tidak baku. Hal tersebut dapat mempengaruhi tingkat akurasi, presisi, dan recall sehingga diperlukannya algoritma tambahan yaitu *Levenshtein Distance*. Oleh karena itu penelitian ini akan melakukan optimasi metode *Multinomial Naïve Bayes* dengan metode *Levenshtein Distance* terhadap data ulasan aplikasi *KAI Access*, sehingga diharapkan dapat meningkatkan nilai akurasi, presisi dan *recall*.

## 2. KAJIAN PUSTAKA

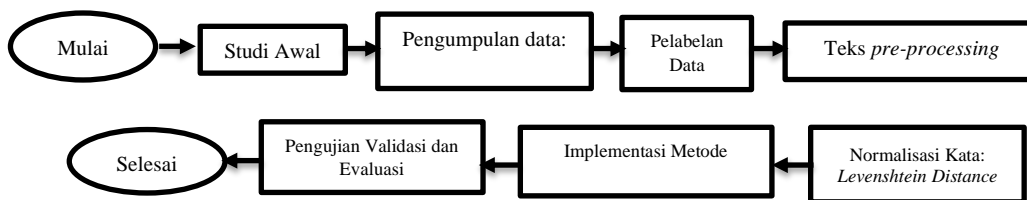
Penelitian oleh Machrus Izunnahdi yang berjudul “*Sentimen Analisis Pada Data Ulasan Aplikasi KAI Access di Google Play Store Menggunakan Metode Multinomial Naïve Bayes*”. Penelitian ini menggunakan pengambilan data ulasan aplikasi *KAI Access* pada *Google Play*. Metode yang dipakai pada penelitian ini yaitu algoritma *Multinomial Naïve Bayes* dengan pengujian data yang mencapai akurasi sebesar 89%, presisi 92% dan recall 85% (Izunnahdi dkk., 2023).

Penelitian oleh Yuyun, Nurul Hidayah dan Supriadi Sahibu yang berjudul “*Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Penanganan Covid-19 Menggunakan Data Twitter*”. Pada penelitian ini menggunakan metode *Multinomial Naïve Bayes* sebanyak 2000 dataset yang bersumber dari media sosial *twitter*, dari hasil pengujian diperoleh bahwa *weighted average* untuk presisi, *recall* dan akurasi diperoleh sebesar 74% (Yuyun dkk., 2021).

Penelitian oleh Sigit Pamungkas yang berjudul “*Klasifikasi Sentiment Tweet Pelanggan Indihome Selama Pandemi Covid -19 Menggunakan Algoritma Multinomial Naïve Bayes*”. Pada penelitian ini data ulasan diperoleh dari *tweet* pelanggan pengguna layanan Indihome. Metode yang digunakan pada penelitian ini yaitu *Multinomial Naïve Bayes* dengan nilai rata-rata akurasi terbaik sebesar 76,07% (Pamungkas & Darmawan, 2022).

## 3. METODE PENELITIAN

Metode penelitian data ulasan aplikasi *KAI Access* dimulai dengan proses *scrapping* data kemudian dilakukan pelabelan data dan *preprocessing text*. Setelah itu data uji akan dilakukan implementasi metode *Levenshtein Distance* dan implementasi metode *Multinomial Naïve Bayes*. Tahap terakhir dilakukan proses pengujian validasi dan evaluasi. Diagram alur secara umum digambarkan sebagai berikut:



Gambar 1. Diagram Alur Penelitian

#### A. Pengumpulan Data dan Pelabelan Data

Jenis data yang akan digunakan pada penelitian ini bersumber dari data ulasan atau *review* pengguna aplikasi *KAI Access* yang ada di *Google Play*. Kemudian teknik pengumpulan data dilakukan dengan teknik *scrapping*. Fitur yang digunakan untuk *scrapping*, pada data terdapat username pengguna, rating dan ulasan komentar. Jumlah populasi atau banyaknya data ulasan komentar *KAI Access* sebanyak 67675, kemudian diterapkan rumus *slovin*  $n = 67675 / (67675 + 0,01^2) = 8712,58$ . Sehingga pada penelitian ini setelah dilakukan perhitungan *slovin* maka akan dilakukan *scrapping* data didapatkan data ulasan aplikasi *KAI Access* sebanyak 8713 data.

Data komentar aplikasi *KAI Access* yang sudah di *scrapping* berikutnya akan dilakukan pelabelan nilai aktual oleh validator dari setiap ulasan pengguna. Pada proses validasi sentimen data ulasan yaitu dibantu menggunakan tenaga ahli. Bantuan tenaga ahli tersebut adalah pegawai PT. KAI.

#### B. Teks Preprocessing

Pada tahap teks *preprocessing* ini berguna untuk mengolah data menjadi lebih terstruktur yang dibagi menjadi beberapa tahapan (Pandunata dkk., 2022). Berikut adalah tahapan-tahapan teks *preprocessing*:

##### 1. *Cleaning*

Tahap *cleaning* akan dilakukan penghapusan karakter yang bukan termasuk alfabet dan mengurangi karakter yang tidak memiliki arti pada analisis sentimen. Contohnya penghapusan seperti @, #, %, \*, dan emoji.

Tabel 1. Proses *Cleaning*

Ulasan	<i>Cleaning</i>
Kenapa jadi ngelag tiap mesen kursi	Kenapa jadi ngelag tiap mesen
Sedih bangeettðŸ˜ˆ-ðŸ˜ˆ-ðŸ˜ˆ-ðŸ˜ˆ-ðŸ˜ˆ-	kursi Sedih bangeett

##### 2. *Case Folding*

Pada tahap *case folding* ini jika terdapat huruf kapital akan diubah menjadi huruf kecil, proses tersebut bertujuan agar data mempunyai suatu format yang sama.

Tabel 2. Proses *Case Folding*

Ulasan	<i>Cleaning</i>
Kenapa jadi ngelag tiap mesen	kenapa jadi ngelag tiap mesen
kursi Sedih bangeett	kursi sedih bangeett

##### 3. *Tokenizing*

Berikutnya adalah *Tokenizing*, pada tahap ini akan dilakukan proses pemisahan dari sebuah kalimat menjadi bagian-bagian per kata.

Tabel 3. Proses *Tokenizing*

Ulasan	<i>Tokenizing</i>
kenapa jadi ngelag tiap mesen kursi sedih bangeett	['kenapa', 'jadi', 'ngelag', 'tiap', 'mesen', 'kursi', 'sedih', 'bangeett']

##### 4. *Stopword Removal*

Tahap *stopword Removal* ini merupakan suatu proses menghapus kata-kata yang tidak memiliki arti atau tidak penting dan memilih kata-kata yang akan digunakan dalam dokumen.

Tabel 4. Proses *Stopword Removal*

Ulasan	<i>Stopword Removal</i>
kenapa jadi ngelag tiap mesen kursi sedih bangeett	['ngelag', 'mesen', 'kursi', 'sedih', 'bangeett']

5. *Stemming*

Tahap *stemming* merupakan tahap untuk mengetahui suatu kata dasar pada sebuah kalimat, jadi jika ada kalimat yang terdapat sebuah imbuhan maka kalimat tersebut akan diubah menjadi bentuk kata dasar.

Tabel 5. Proses *Stemming*

Ulasan	<i>Stemming</i>
['jadi', 'ngelag', 'tiap', 'mesen', 'kursi', 'sedih', 'bangeett']	ngelag mesen kursi sedih bangeett

C. *Levenshtein Distance*

Metode *Levenshtein Distance* merupakan salah satu penerapan metode yang dapat digunakan untuk memperbaiki terjadinya kesalahan ejaan pada kata, manfaat dari metode ini adalah untuk menghitung dan mengetahui suatu jarak terdekat dari duah *string* melalui proses perubahan karakter (Rosmala & Risyad, 2018), penambahan karakter, dan penghapusan dari suatu karakter sampai kedua *string* tersebut cocok satu sama lain.

Pada tahap implementasi menggunakan metode *Levenshtein Distance* dilakukan perhitungan untuk memperoleh jarak terdekat antara dua buah string. Metode *Levenshtein Distance* digunakan untuk menganalisa perbandingan kesamaan atau perbedaan dari dua buah *string*. Berikut adalah alur implementasi metode *Levenshtein Distance*:

1. Menyiapkan dan membangun kamus untuk kata dasar yang bersumber dari KBBI sebanyak 29.932 kata.
2. Melakukan identifikasi setiap kata pada ulasan aplikasi lalu dibandingkan dengan kamus kata dasar KBBI menggunakan metode *Levenshtein Distance*.
3. Melakukan perhitungan matriks untuk mencari jarak terdekat pada tiap-tiap kata menggunakan rumus *Levenshtein Distance* yaitu:

$$\begin{aligned}
 D(s, t) &= \min D(s - 1, t) + 1 \\
 D(s, t) &= \min D(s, t - 1) + 1 \\
 D(s, t) &= \min D(s - 1, t - 1) + 1, s_j \neq t_i \\
 D(s, t) &= \min D(s - 1, t - 1) + 1, s_j = t_i
 \end{aligned}
 \tag{1}$$

Keterangan:

- s* : string pertama atau sumber
- b* : kata kedua atau target
- s(j)* : karakter dari *string* sumber ke-*j*
- t(i)* : karakter dari *string* sumber ke-*i*
- D* : jarak edit *Levenshtein Distanc*

4. Apabila pada kata yang dibandingkan dengan kamus kata dasar tidak ditemukan maka akan dikembalikan ke kata awal atau kata semula.

Contoh perhitungan *Levenshtein Distance*:

Tabel 6. Contoh Perhitungan *Levenshtein Distance*

String Target	String Number					
	(i)	"m"	"e"	"s"	"n"	
(j)	0	1	2	3	4	5
"p"	1	1	2	3	4	5
"e"	2	2	1	2	3	4
"s"	3	3	2	1	2	3
"a"	4	4	3	2	2	3
"n"	5	5	4	3	3	2

Contoh hasil perhitungan *Levenshtein Distance*:

Tabel 7. Contoh hasil perhitungan *Levenshtein Distance*

String 1	String 2	Jarak <i>Levenshtein Distance</i>	Waktu Eksekusi (s)
“ya”	“iya”	1	0.91 seconds
“mesen”	“pesan”	2	3.67 seconds

#### D. Implementasi *Multinomial Naïve Bayes*

Setelah melakukan tahapan normalisasi *Levenshtein Distance* dan *preprocessing text*, akan dilakukan tahapan berikutnya yaitu pembobotan kata dengan metode ekstraksi fitur TF-IDF untuk mengubah bentuk teks yang terdapat pada komentar menjadi numerik. *Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan proses pembobotan pada suatu kata yang berguna untuk mentransformasi data atau kata dari tekstual menjadi bentuk data numerik (Septian dkk., 2019). Data yang sudah menjadi bentuk numerik selanjutnya akan dilakukan perhitungan klasifikasi sentimen dengan metode *Multinomial Naïve Bayes*.

*Multinomial Naïve Bayes* adalah metode memperhitungkan jumlah kemunculan suatu kata dalam dokumen, dengan asumsi dokumen memiliki beberapa kejadian pada kata dengan panjang yang tidak tergantung berdasarkan kelasnya dalam dokumen. Metode *Multinomial Naïve Bayes* sangatlah cocok digunakan untuk menghitung dokumen atau data teks dengan jumlah kata besar dan *multiclass* (Haditira dkk., 2022).

$$P(C) = \frac{\text{count}(c)+K}{N+K \cdot |\text{classes}|} \quad (2)$$

Keterangan:

- P : probabilitas dari variable c.
- Count : jumlah suatu kemunculan dari sampel c.
- K : nilai parameter.
- N : jumlah total dari kejadian sampel c.
- |\text{classes}| : jumlah seluruh kelas yang ada pada sampel.

## 4. HASIL DAN PEMBAHASAN

Skenario uji coba yang diterapkan pada penelitian ini menggunakan *K-Fold Cross Validation* dengan nilai K yaitu 2,3,4,5,6,7,8,9,10.. Pengujian pada penelitian ini dilakukan sebanyak dua kali proses uji coba, yaitu yang pertama menggunakan metode *Multinomial Naïve Bayes* dan yang kedua yaitu menggunakan metode *Multinomial Naïve Bayes* yang dioptimasi menggunakan metode *Levenshtein Distance*.




Dari keseluruhan proses pengujian klasifikasi data dengan menggunakan metode *Multinomial Naïve Bayes* yang dioptimasi dengan metode *Levenshtein Distance* menggunakan *K-Fold* 2,3,4,5,6,7,8,9,10 maka hasilnya yaitu sebagai berikut:

Tabel 8. Hasil Pengujian dan Validasi

<i>K-Fold</i>	Pembagian ke-	<i>Multinomial naïve bayes</i>			<i>Multinomial naïve bayes dan levenshtein distance</i>			Data Training
		<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	
2	1	86%	59%	45%	85%	58%	44%	4375
	2	74%	56%	38%	74%	55%	39%	4376
3	1	86%	59%	44%	86%	58%	44%	2905
	2	81%	55%	41%	81%	55%	42%	2904
	3	73%	56%	40%	73%	56%	40%	2904
4	1	87%	59%	45%	87%	58%	44%	2179
	2	84%	57%	44%	84%	58%	43%	2178
	3	80%	55%	42%	80%	55%	43%	2178
	4	72%	55%	41%	71%	54%	41%	2178
5	1	88%	59%	45%	87%	57%	44%	1743
	2	85%	57%	43%	85%	57%	43%	1743
	3	79%	53%	42%	79%	54%	42%	1743
	4	80%	58%	43%	80%	57%	44%	1742
	5	71%	55%	42%	72%	55%	42%	1742
6	1	88%	58%	45%	88%	58%	45%	1453

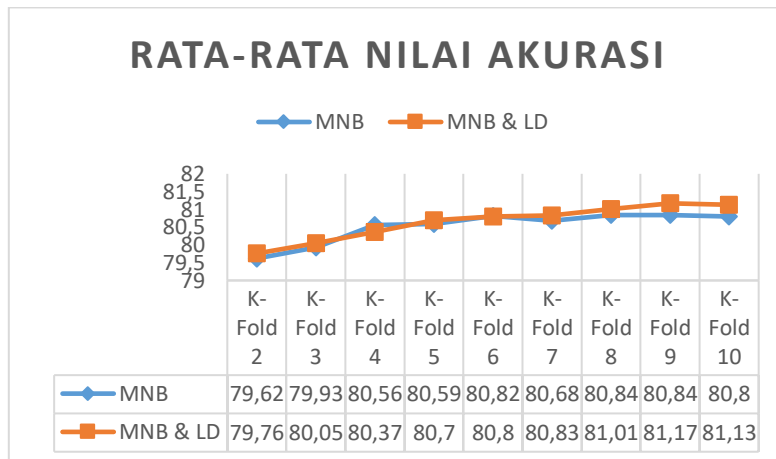
K-Fold	Pembagian ke-	Multinomial naïve bayes			Multinomial naïve bayes dan levenshtein distance			Data Training
		Accuracy	Precision	Recall	Accuracy	Precision	Recall	
7	2	85%	60%	44%	85%	60%	44%	1452
	3	84%	56%	44%	84%	56%	44%	1452
	4	79%	52%	41%	79%	52%	41%	1452
	5	79%	58%	44%	79%	58%	44%	1452
	6	71%	55%	42%	71%	55%	42%	1452
	1	89%	58%	45%	89%	57%	44%	1245
8	2	83%	58%	43%	83%	58%	43%	1245
	3	86%	57%	43%	86%	58%	43%	1245
	4	79%	53%	42%	80%	54%	43%	1245
	5	80%	55%	42%	80%	57%	45%	1245
	6	77%	57%	44%	77%	55%	44%	1244
	7	70%	55%	42%	70%	54%	42%	1244
	1	90%	57%	45%	90%	55%	44%	1090
9	2	84%	59%	44%	84%	59%	44%	1089
	3	84%	56%	43%	84%	56%	43%	1089
	4	82%	56%	43%	83%	59%	45%	1089
	5	78%	51%	41%	78%	50%	41%	1089
	6	82%	58%	44%	82%	59%	45%	1089
	7	76%	56%	44%	75%	54%	43%	1089
	8	71%	55%	43%	72%	56%	44%	1089
	1	91%	56%	45%	91%	54%	43%	969
10	2	83%	59%	44%	84%	60%	46%	968
	3	84%	59%	44%	84%	57%	43%	968
	4	85%	56%	43%	86%	58%	44%	968
	5	79%	52%	41%	80%	53%	43%	968
	6	78%	52%	41%	79%	53%	42%	968
	7	80%	59%	45%	80%	59%	45%	968
	8	75%	55%	45%	76%	55%	45%	968
	9	71%	55%	43%	73%	56%	45%	968
	1	92%	54%	43%	91%	52%	42%	872
	2	83%	59%	44%	83%	59%	45%	872
3	83%	59%	44%	82%	56%	42%	872	
4	87%	54%	43%	88%	57%	45%	871	
5	81%	56%	43%	82%	58%	45%	871	
6	77%	50%	41%	77%	49%	41%	871	
7	82%	58%	42%	83%	57%	43%	871	
8	78%	57%	44%	78%	57%	45%	871	
9	74%	55%	44%	75%	55%	45%	871	
10	71%	55%	43%	73%	56%	45%	871	

Keterangan:

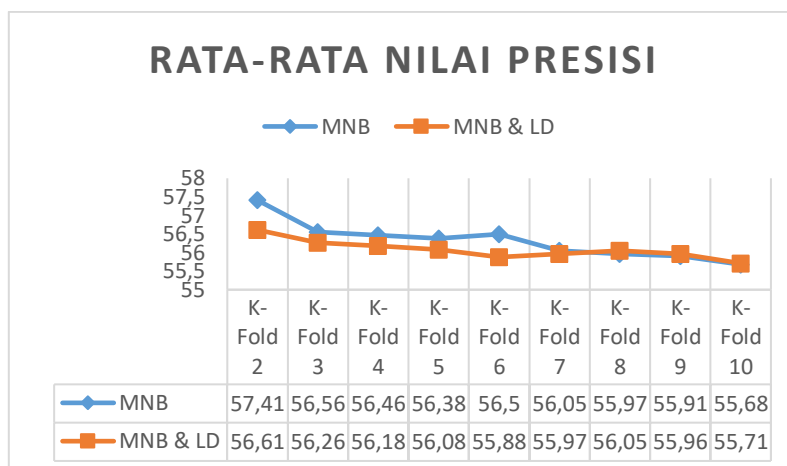
-  : Nilai tertinggi pengujian menggunakan metode *Multinomial Naïve Bayes*
-  : Nilai tertinggi peningkatan akurasi menggunakan kombinasi dua metode *Multinomial Naïve Bayes* dan *Levenshtein Distance*
-  : Nilai tertinggi pengujian menggunakan metode *Levenshtein Distance*

Pada klasifikasi ulasan pengguna aplikasi KAI Access dengan jumlah 8713 data menggunakan metode *Multinomial Naïve Bayes* memperoleh nilai akurasi tertinggi sebesar 92% lalu nilai presisi sebesar 60% dan nilai *recall* sebesar 45%. Kemudian ketika menggunakan metode *Multinomial Naïve Bayes* yang dioptimasi dengan metode *Levenshtein Distance* memperoleh nilai akurasi tertinggi sebesar 91% lalu nilai presisi sebesar 60% dan nilai *recall* sebesar 46%.

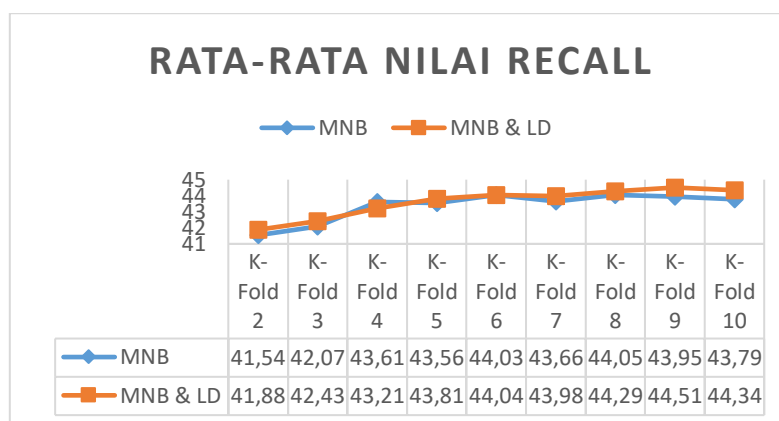
Dari hasil pengujian evaluasi dan validasi menggunakan *K-Fold* 2,3,4,5,6,7,8,9,10 menggunakan metode *Multinomial Naïve Bayes* saja dan menggunakan kombinasi *Multinomial Naïve Bayes* dengan optimasi *Levenshtein Distance*, diperoleh hasil rangkuman rata-rata sebagai berikut:



Gambar 2. Rata-Rata Perbandingan Nilai Akurasi



Gambar 3. Rata-Rata Perbandingan Nilai Presisi



Gambar 4. Rata-Rata Perbandingan Nilai Recall

Berdasarkan rata rata perbandingan akurasi, presisi dan *recall* dapat dianalisa bahwa validasi *K-Fold Cross* harus diperhatikan secara detail dan hati-hati ketika menentukan nilai *K-Fold* yang akan dipilih. Dari hasil rata-rata perhitungan *K-Fold* dengan metode *Multinomial Naïve Bayes* dengan dan *Levenshtein Distance* dapat meningkatkan nilai akurasi, presisi dan *recall* yang terjadi pada uji coba

*K-Fold* ke 7,8,9,10 hal ini dikarenakan dengan menggunakan *K-Fold* yang semakin banyak setiap nilai *K* memiliki kombinasi yang representatif dari data uji.

## 5. KESIMPULAN

Berdasarkan proses penelitian yang sudah dilakukan terhadap ulasan aplikasi KAI *Access* menggunakan metode *Multinomial Naïve Bayes* dengan dioptimasi menggunakan metode *Levenshtein Distance* memperoleh hasil akhir dan dapat ditarik kesimpulan sebagai berikut:

- 1) Optimasi *Levenshtein Distance* terhadap algoritma *Multinomial Naïve Bayes* dapat meningkatkan nilai akurasi, presisi dan *recall*. Yaitu terjadi pada *K-Fold* 9 di percobaan ke 2 yaitu yang awalnya menggunakan *Multinomial Naïve Bayes* sebesar mendapatkan nilai akurasi 83%, presisi 59% dan *recall* 44% kemudian meningkat ketika menggunakan kombinasi metode *Levenshtein Distance* sebesar akurasi 84%, presisi 60% dan *recall* 46%.
- 2) Dari pengujian *K-Fold* 2,3,4,5,6,7,8,9,10 menggunakan metode *Multinomial Naïve Bayes* diperoleh hasil pengukuran tertinggi tingkat akurasi yang diperoleh sebesar 92%, tingkat presisi sebesar 60% dan tingkat *recall* sebesar 45%. Nilai tertinggi ketika menggunakan metode *Multinomial Naïve Bayes* dan *Levenshtein Distance* diperoleh hasil pengukuran tertinggi yaitu tingkat akurasi yang diperoleh sebesar 91%, tingkat presisi sebesar 60% dan tingkat *recall* sebesar 46%.
- 3) Metode *Levenshtein Distance* pada prosesnya tidak mempertimbangkan tingkat kepentingan kata, jadi setiap perbedaan karakter pada suatu kata dianggap memiliki makna yang sama pentingnya. Jadi proses pengujian tidak memperhatikan kata yang mungkin memiliki kontribusi yang bermakna atau lebih besar dalam menentukan analisis sentimen. Hal tersebut dapat berpengaruh sehingga mengakibatkan naik turunnya tingkat akurasi yang diperoleh.

Pada penelitian ini terdapat beberapa kekurangan sehingga tidak sempurna. Oleh karena itu penulis berharap terdapat pengembangan dari peneliti berikutnya.

- 1) Menggunakan metode normalisasi selain *Levenshtein Distance* pada analisis sentimen untuk memperoleh hasil yang maksimal pada proses normalisasi kata dan dapat memperoleh tingkat akurasi yang lebih maksimal.
- 2) Dapat menggunakan metode klasifikasi lain dan terbaru untuk membandingkan dan menemukan hasil yang lebih baik.

## 6. DAFTAR PUSTAKA

- Haditira, R., Murdiansyah, D. T., & Astuti, W. (2022). Analisis Sentimen Pada Steam Review Menggunakan Metode *Multinomial Naïve Bayes* dengan Seleksi Fitur Gini Index Text. *e-Proceeding of Engineering*, 9(3), 1793–1799.
- Izunnahdi, M., Aburrahman, G., & Eko Wardoyo, A. (2023). Sentimen Analisis Pada Data Ulasan Aplikasi KAI Access Di Google PlayStore Menggunakan Metode Multinomial Naive Bayes Sentiment Analysis on KAI Access Application Review Data on Google PlayStore Using Multinomial Naive Bayes Method. *Jurnal Smart Teknologi*, 4(2), 192–198. <http://jurnal.unmuhjember.ac.id/index.php/JST>
- Lestari, S., & Saepudin, S. (2021). Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naive Bayes. *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, 163–170.
- Pamungkas, S., & Darmawan, J. B. B. (2022). Klasifikasi Sentiment Tweet Pelanggan IndiHome Selama Pandemi Covid-19 Menggunakan Algoritma Multinomial Naive Bayes. *SNESTIK Seminar Nasional Teknik Elektro*, 339–344.
- Pandunata, P., Ananta, C. K., & Nurdiansyah, Y. (2022). Analisis Sentimen Opini Publik Terhadap Pekan Olahraga Nasional Pada Instagram Menggunakan Metode Naïve Bayes Classifier. *INFORMAL: Informatics Journal*, 7(2), 146–156.



- Rosmala, D., & Risyad, Z. M. (2018). Algoritma *Levenshtein Distance* dalam Aplikasi Pencarian isu di Kota Bandung pada Twitter. *MIND Journal*, 2(2), 1–12.
- Septian, J. A., Fachrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *Journal of Intelligent System and Computation*, 1(1), 43–49.
- Suwondo, A., Sarana, & Marjan, F. I. (2017). Analisis Pengaruh E-Kepuasan pelanggan terhadap E-Loyalitas pelanggan KAI Access Berdasarkan E-SERVQUAL pada PT Kereta Api Indonesia (Persero) DAOP IV Semarang. *Sentrinov*, 3, 338–360.
- Yuyun, Hidayah, N., & Sahibu, S. (2021). Algoritma *Multinomial Naïve Bayes* Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(4), 820–826.