

Contents list available at www.jurnal.unimed.ac.id

CESS
(Journal of Computing Engineering, System and Science)

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



Analisis Perbandingan Multinomial Naïve Bayes dan Adaboost dalam Mengklasifikasikan Sentimen Terkait Pinjaman Online

Comparative Analysis of Multinomial Naïve Bayes and Adaboost in Classifying Sentiment Related to Online Loans

Yoga Pangestu^{1*}, Muhammad Basri²

^{1,2} Universitas Muhammadiyah Sumatera Utara

Jl. Kapt. Mukhtar Basri No. 3 Medan Sumatera Utara

email: ¹yogapangestuuu23@gmail.com, ²muhammadbasri@umsu.ac.id

ABSTRAK

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma *Multinomial Naive Bayes* (MNB) dan AdaBoost dalam mengklasifikasikan sentimen masyarakat terkait pinjaman online. Data yang digunakan dalam penelitian ini adalah komentar pengguna di media sosial twitter terkait pinjaman online. Metode yang digunakan untuk mengolah data adalah SEMMA (*Sample, explore, modify, model, acces*), tahapan tersebut mencakup tahap *preprocessing* data pemodelan dan evaluasi. Sentiment diklasifikasikan kedalam kelas positif, negatif dan netral dengan menggunakan kamus lexicon base bahasa indonesia. Model yang dibangun menggunakan Algoritma MNB dan AdaBoost untuk dibandingkan performanya. Hasil evaluasi model menunjukkan menunjukkan bahwa algoritma AdaBoost memiliki kinerja yang lebih baik dibandingkan dengan algoritma MNB dalam mengklasifikasikan sentimen masyarakat terkait pinjaman online. Hal ini dibuktikan dengan nilai akurasi dari algoritma AdaBoost sebesar 76%, sedangkan akurasi dari algoritma MNB sebesar 71%.

Kata Kunci: Analisis sentimen; Pinjaman online; Multinomial Naive Bayes; AdaBoost, klasifikasi; Data Mining.

ABSTRACT

This research aims to analyze and compare the performance of the Multinomial Naive Bayes (MNB) and AdaBoost algorithms in classifying public sentiment regarding online loans. The data used in this research are user comments on Twitter social media regarding online loans. The method used to process data is SEMMA (*Sample, explore, modify, model, access*), this stage includes the preprocessing data modeling and evaluation stages. Sentiment is classified into positive, negative and neutral classes using the Indonesian lexicon base dictionary. The

*Penulis Korespondensi:

email: yogapangestuuu23@gmail.com

model built uses the MNB and AdaBoost algorithms to compare their performance. The model evaluation results show that the AdaBoost algorithm has better performance compared to the MNB algorithm in classifying public sentiment regarding online loans. This is proven by the accuracy value of the AdaBoost algorithm being 76%, while the accuracy of the MNB algorithm is 71%.

Keywords: *Sentiment analysis; online loans; Multinomial Naive Bayes; AdaBoost; classification*

1. PENDAHULUAN

Pinjaman online merupakan perkembangan teknologi dari konsep peminjaman uang yang kini bisa melalui online atau menggunakan teknologi untuk memudahkan transaksi. Pinjaman online telah menjadi salah satu solusi finansial yang populer bagi masyarakat di era digital saat ini. Pinjaman online menawarkan kemudahan dan kecepatan dalam mendapatkan pinjaman, sehingga menjadi solusi bagi masyarakat yang membutuhkan dana cepat. Meskipun memberikan akses cepat dan mudah untuk mendapatkan dana, masih banyak penyedia layanan pinjaman online yang belum memiliki regulasi yang memadai sehingga potensi risiko bagi para peminjam menjadi lebih tinggi. Fenomena ini juga menimbulkan perhatian atas beragam isu terkait, termasuk kesulitan pembayaran, suku bunga yang tinggi, dan praktik peminjaman yang tidak bertanggung jawab[1]. Hal ini diperparah dengan maraknya kasus penipuan dan tindakan yang tidak etis oleh beberapa pemberi pinjaman online.

Oleh karena itu timbulah berbagai macam persepsi masyarakat terhadap penggunaan pinjaman online. Untuk mengetahui persepsi masyarakat tersebut perlu adanya analisa sentiment, analisa sentimen digunakan untuk mengidentifikasi pola dari opini dan persepsi yang beredar sehingga diketahui fenomena tersebut berdampak positif atau negatif bagi masyarakat. sentiment masyarakat tersebut dapat dilihat dari media sosial twitter. Twitter menjadi media sosial yang penggunanya paling aktif terhadap suatu isu sehingga cocok untuk sumber data[2].

Analisa sentiment masuk kedalam cabang ilmu data mining yaitu *text mining*, *text mining* sendiri adalah proses analisa data berupa text. *Text mining* merupakan teknologi yang digunakan untuk menganalisis data tak terstruktur berbentuk teks. Dalam analisis *text mining* terdapat dua fase utama yaitu *preprocessing* dan integrasi dari data tak terstruktur, dan analisis statistik data yang telah dilakukan *preprocessing* untuk mengekstraksi konten dari yang terdapat dalam teks. Dengan kata lain text mining mengubah data tak terstruktur menjadi data terstruktur[3]. Pada penelitian ini pengolahan text mining digunakan metode text klasifikasi yang mana nantinya akan mengklasifikasikan sentiment dari media sosial twitter kedalam tiga kelas yaitu positif negatif dan netral. Klasifikasi text tersebut dibantu dengan menggunakan kamus lexicon bahasa indonesia, Kamus lexicon tersebut didapat dari penelitian yang dilakukan oleh Koto & Rahmaningtyas[4] yang mengumpulkan kata-kata positif dan negatif berbahasa Indonesia yang ada pada blog dan memberinya bobot nilai plus dan minus. Kamus tersebut juga telah diuji dalam jurnal yang ditulis oleh Azhar[5] tentang identifikasi opini pada tweet berbahasa Indonesia dan hasilnya akurat.

Untuk menganalisa sentiment tersebut digunakan algoritma Multinomial Naive Bayes dan Adaboost. Kedua algoritma tersebut nantinya akan dibandingkan performanya dengan

melihat nilai akurasi, precision, recall dan f1-score dari masing-masing model algoritma. Algoritma Multinomial Naïve Bayes Classifier adalah modifikasi dari algoritma Naïve Bayes Classifier yang terbukti efektif dalam klasifikasi sentimen dan memiliki kinerja lebih baik dari algoritma klasifikasi lainnya[6]. Pada penelitian yang dilakukan oleh Zaki Hariansyah [3], menghasilkan skor akurasi sebesar 93%, precision 90%, recall 93%, dan f1-score 91% dengan rata-rata skor sebesar 92%. Hal tersebut diperkuat dengan penelitian yang dilakukan oleh Zelin Gaa Ngilo & Nuryuliani Nuryulian[7], dengan menggunakan algoritma Multinomial Naïve Bayes diperoleh tingkat akurasi pelatihan sebesar 91.50% dan tingkat akurasi validasi sebesar 85.35% dengan menggunakan 2211 data uji dan 553 data validasi yang bersumber dari twitter.

Adaboost adalah singkatan dari “Adaptive Boosting” merupakan algoritma yang cukup populer untuk melakukan klasifikasi. Algoritma Adaboost memiliki performa yang lebih baik jika dibandingkan dengan algoritma Boosting lainnya seperti Gradien Boosting dan XGBoost . Dari penelitian yang dilakukan Syah dengan menggunakan algoritma AdaBoost didapat nilai Accuracy 80.21, Precision 85.01, Recall 73.36 dan AUC 0.861[8]. Penelitian dengan menggunakan algoritma Adaboost juga dilakukan oleh Wahyu dengan 560 tweet[9]. Ini dibagi menjadi dua, 500 untuk data latihan dan 60 untuk data uji, menghasilkan algoritma AdaBoost memiliki kemampuan untuk mengkategorikan tweet kenaikan BBM Pertamina ke dalam kelas bersentimen positif atau negatif dengan akurasi sebesar 86,8%.

Atas dasar maraknya stigma masyarakat terkait isu pinjaman online maka dilakukan penelitian dengan cara melakukan analisa sentimen. Analisis sentimen atas kasus pinjaman online dapat menjadi kunci untuk memahami persepsi, pendapat, serta perasaan yang berkembang di kalangan masyarakat terkait masalah ini. Dengan adanya penelitian ini juga akan melihat kinerja algoritma mana yang lebih baik antara Multinomial Naïve Bayes dan AdaBoost untuk mengklasifikasikan sentiment masyarakat terkait isu pinjaman online.

2. TINJAUAN TEORI

2.1. Data Mining

Data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan. Data mining merupakan teknik yang dapat membantu memprediksi fenomena suatu objek yang diteliti dengan terlebih dahulu mengetahui atau tidak mengetahui suatu kelas suatu objek tertentu melalui masukan dari data besar[10]. Sedangkan menurut Munarsih[11], Data mining merupakan bidang ilmu yang digunakan untuk menangani masalah pengambilan informasi dari database yang besar dengan menggabungkan teknik dari statistik, pembelajaran mesin, visualisasi data, pengenalan pola, dan database.

Salah satu metode penyelesaian masalah yang ada dalam data mining adalah text mining. Text mining adalah metode mengekstrak informasi dari sekumpulan teks untuk menemukan pola dan informasi yang valid dari sebuah teks. Text mining mengacu pada ekstraksi informasi dan pola yang implisit, sebelumnya tidak diketahui, dan berpotensi berharga secara otomatis atau semi-otomatis dari data tekstual tidak terstruktur yang sangat besar, seperti teks bahasa alami. Data mining dan text mining berbeda berdasarkan jenis data yang ditanganinya. Sementara data mining menangani data terstruktur yang berasal dari sistem, seperti database, spreadsheet, ERP, CRM, dan aplikasi akuntansi, text mining menangani data tidak terstruktur yang ditemukan di dokumen, email, media sosial, dan web.

Dengan demikian, perbedaan antara data mining dan text mining adalah bahwa dalam text mining polanya diekstraksi dari teks bahasa alami dan bukan dari database fakta yang terstruktur[12].

2.2. Analisa Sentiment

Analisis sentiment adalah bidang data mining yang biasa digunakan untuk menganalisis data teks berupa opini dengan cara menemukan polarisasi dari data tersebut dan kemudian menghasilkan informasi yang bernilai positif, negatif, atau netral. Metode analisis sentimen merupakan suatu metode untuk menganalisis informasi yang diperoleh dari internet dengan sedemikian rupa sehingga memungkinkan untuk mengetahui polaritas informasi tersebut. Analisis sentimen dapat digunakan untuk mengumpulkan polaritas opini yang ada untuk digunakan dalam memprediksi sentimen publik[9].

Analisis sentimen melibatkan penambangan data, yang menganalisis, memproses, dan mengekstrak data tekstual dari entitas seperti layanan, produk, manusia, fenomena, dan topik. Proses analisis dapat mencakup teks ulasan, forum, tweet, blog, dll. Text mining adalah proses menganalisis data dalam bentuk teks.

2.3. Multinomial Naïve Bayes

Multinomial Naïve Bayes merupakan percabangan dari algoritma Naïve Bayes yang digunakan untuk klasifikasi. Metode Naïve Bayes terbukti dapat memberikan hasil yang cukup memuaskan ketika digunakan untuk klasifikasi teks[13]. Seperti namanya, metode ini mengasumsikan bahwa setiap atribut dari data tidak bergantung satu sama lain. Pada dasarnya, asumsi bahwa setiap kata tidak bergantung satu dengan yang lain pada metode Naïve Bayes ini berlawanan dengan keadaan sebenarnya. Hal ini dikarenakan suatu dokumen atau teks perlu memiliki kata yang saling berhubungan agar dokumen tersebut memiliki makna. Akan tetapi, metode ini terbukti mampu memberikan hasil yang cukup memuaskan apabila diterapkan di bidang klasifikasi teks.

Algoritma Multinomial naive bayes merupakan salah satu metode khusus dari naive bayes sebagai metode text mining dalam proses klasifikasi teks menggunakan probabilitas kelas dalam dokumen. Peluang untuk memperkirakan prior probability $P(c)$ atau peluang kemunculan suatu kelas pada data latih dilakukan menggunakan persamaan [14]:

$$P(c) = \frac{Nc}{Ndoc} \quad (1)$$

Keterangan:

c : Kategori atau kelas

doc : Dokumen

Nc : Banyaknya kategori c pada dokumen latih

$Ndoc$: Banyaknya keseluruhan dokumen latih yang digunakan

Perhitungan selanjutnya dari probabilitas bahwa setiap kata termasuk dalam kategori atau kelas tertentu dapat dilakukan dengan menggunakan persamaan:

$$P(w_i, c) = \frac{count(w_i, c) + 1}{\sum_w count(w, c) + |V|} \quad (2)$$

Keterangan:

w_i : Kata ke- i dalam seluruh dokumen yang berkategori c

- (w_i) : Jumlah kata tertentu yang muncul dalam suatu kategori atau kelas
 $\sum w(w_i)$: Jumlah seluruh kata pada kelas
 $|V|$: merupakan jumlah seluruh kata unik pada kelas

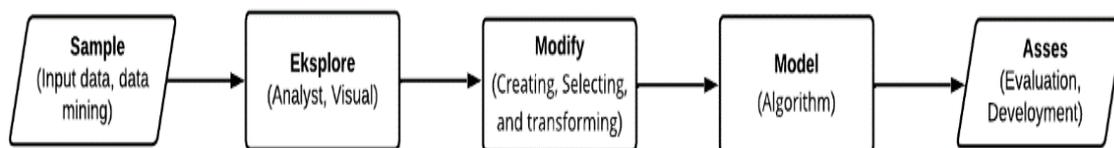
2.4. Adaboost

Adaboost atau kepanjangan dari Adaptive Boosting merupakan percabangan dari algoritma Boosting diantaranya adalah XGBoost, Gradien Boost dan Adaboost. Algoritma Adaboost (Adaptive Boosting) adalah algoritma pembelajaran mesin yang digunakan untuk meningkatkan kinerja model prediksi dengan menggabungkan beberapa model prediksi sederhana menjadi satu model yang lebih kompleks. Algoritma ini bekerja dengan memberikan bobot yang berbeda pada setiap sampel data, sehingga sampel yang sulit diprediksi akan diberikan bobot yang lebih tinggi. Kemudian, model prediksi sederhana akan dibuat untuk setiap iterasi dengan memperhatikan bobot pada setiap sampel data. Model-model tersebut kemudian digabungkan menjadi satu model yang lebih kompleks dengan menggunakan teknik voting[9].

Algoritma AdaBoost memiliki 3 tahap. Tahap pertama memulai pendistribusian bobot data pelatihan. Jika dimiliki nilai sampel N , pada awalnya setiap sampel pelatihan akan diberikan bobot yang sama. Tahap kedua yaitu melatih pengelompokan dasar. Dalam proses yang spesifik, jika sebuah sampel telah diklasifikasikan dengan akurat, maka bobot akan dikurangi pada set yang akan dibangun berikutnya. Jika sampel diklasifikasikan belum akurat, maka bobot akan ditingkatkan pada set yang akan dibangun berikutnya. Set berikutnya digunakan untuk melatih pengelompokan berikutnya. Pada tahap terakhir yaitu menggabungkan weak classifiers yang diperoleh dari setiap pelatihan menjadi strong classifier. Setelah melatih tiap golongan yang lemah, meningkatkan bobot pada weak classifier dengan klasifikasi error rate yang kecil akan berperan lebih baik pada klasifikasi akhir. Dengan kata lain, weak classifier dengan nilai error rate yang kecil akan memiliki bobot yang besar pada pengklasifikasian akhir, begitu pula sebaliknya[15].

3. METODE

Metode pengolahan data menggunakan metode SEMMA (sample, explore, modify, model acces), tahapan tersebut telah mencakup pengumpulan data, preprocessing data, modeling dan evaluasi model.



Gambar 1. Tahapan Metode SEMMA

3.1. Sample

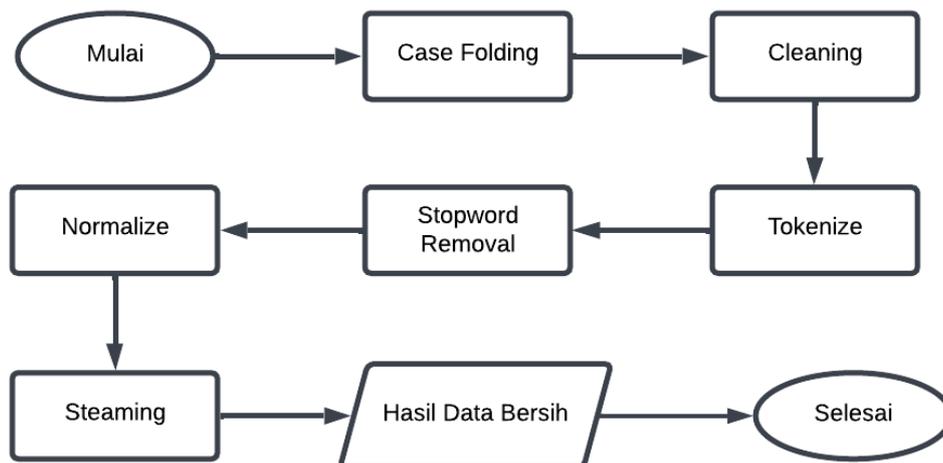
Tahapan sample adalah pengumpulan data yang diambil dari media social twitter dengan metode scraping dibantu dengan API twitter dan autentikasi nomor akun pengguna. Proses scraping data menggunakan tools Google Collab dengan query “pinjol”.

3.2. Explore

Pada tahap ini data dideskripsikan dan dieksplorasi untuk tren yang tak terduga dan anomali dalam rangka untuk mendapatkan pemahaman data dan ide-ide. Pada tahapan explore juga dilakukan penyeleksian atribut yang tidak digunakan dalam analisa data (*data cleaning*), hal ini berguna supaya modem bekerja lebih maksimal. Pada tahap ini juga dilihat kata-kata yang sering muncul dalam data untuk melihat hasil kata yang sering muncul sebelum dan sesudah dilakukan tahapan *text preprocessing*.

3.3. Modify

Modify adalah tahapan *preprocessing* data yang mana data akan diolah untuk mendapatkan data bersih yang siap dimasukan kedalam model algoritma. Preprocessing data dimulai dengan case folding, cleaning, tokenize, stopword removal, normalize dan stemming. Tahapan case folding adalah untuk mengubah semua karakter yang ada pada data menjadi huruf kecil, cleaning adalah tahapan untuk membersihkan data dari kolom yang kosong, tokenize adalah untuk memecah data menjadi bentuk token atau kata perkata. Normalisasi Merupakan proses untuk mengubah teks pada sebuah dokumen dari kata yang tidak tepat atau kata singkatan bahasa menjadi kata yang memiliki arti. Stopwords Removal adalah tahap untuk mengambil kata yang dianggap penting dari hasil normalisasi atau membuang kata yang dianggap tidak memiliki arti penting. Stemming adalah proses pengolahan teks yang berfungsi untuk memotong imbuhan dari setiap kata dan membuat setiap kata tersebut menjadi kata dasar.



Gambar 2. Tahapan Preprocessing Data

Data yang telah siap diolah tersebut selanjutnya akan diberikan label kelas positif, negatif dan netral dengan acuan kamus lexicon bahasa indonesia.

Table 1. Kamus lexicon base Indonesia positif dan negatif

Kata	Bobot
hai	3
merekam	2
ekstensif	3
paripurna	1
Detail	2
pernik	3
belas	2
welas	4

Kata	Bobot
putus	-2
gelebah	-2
tersentuh	-1
isak	-5
larat	-3
nelangsa	-3
tidak segan	-2

Kata dalam kumpulan dataset akan dibandingkan dengan dokumen kamus lexicon. Perbandingan kata tersebut nantinya diberikan nilai atau score jika kata pada kalimat atau dokumen yang ada pada data, terdapat kesamaan dengan kata yang ada didalam kamus lexicon. Jumlah nilai atau score ini menentukan komentar tersebut termasuk didalam label positif atau negatif, jika nilai score komentar tersebut bernilai 0 maka komentar tersebut tergolong kelas netral. Kamus lexicon yang digunakan bersumber dari yaitu Indonesian Sentiment[4]. Terdapat kurang lebih 10.250 kata yang diberi nilai dari -5 hingga +5.

3.4. Model

Pembangunan model algoritma menggunakan multinomial naïve bayes dan adaboost, keduanya akan dibandingkan performa dari masing-masing model dalam melakukan analisa data. Data yang telah dilabelkan dengan kamus lexicon bahasa indonesia kan menjadi data aktual dan model algoritma akan membuat prediksi klasifikasi data berdasarkan data aktual yang ada. Performa dari model algoritma akan dilihat dari nilai akurasi, presisi, recall dan f1 score dari masing-masing model. Model yang baik akan menghasilkan nilai yang lebih besar.

3.5. Asses

Evaluasi model dapat dilihat dengan menggunakan confusion matrix dan perhitungan nilai akurasi dan yang lainnya akan menggunakan rumus sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (3)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

$$F1\text{-score} = \frac{\text{presisi} \times \text{recall}}{\text{presisi} + \text{recall}} \times 2 \quad (6)$$

4. HASIL DAN PEMBAHASAN

Scraping data twitter mendapatkan data sebanyak 2.896 data. Data yang diambil berupa tweet dari 2023-12-01 sampai dengan 2024-01-31. Data mentah tersebut dilakuakn proses preprocessing untuk menghasilkan data bersih yang siap untuk diolah. Tahapan preprocessing meliputi case folding, cleaning, tokenize, stopwords removal, normalize dan stemming. Dari proses tersebut didapatkan data bersih sebanyak 2562. Kemudian dilakuakn proses labeling data untuk menentukan kelas positif, negatif dan netral dari data. Dari hasil labeling data dengan kamus lexicon didapatkan bahwa 1.591 data masuk kedalam kelas negatif, 509 data masuk kedalam kelas positif dan 462 masuk kedalam kelas netral.

```

# Membuat class untuk membaca kamus lexicon positif dan negatif
lexicon_positive = dict()
with open("/content/kamus_Positif.csv", "r") as csvfile:
    reader = csv.reader(csvfile, delimiter=",")
    for row in reader:
        lexicon_positive[row[0]] = int(row[1])

lexicon_negative = dict()
with open("/content/kamus_Negatif.csv", "r") as csvfile:
    reader = csv.reader(csvfile, delimiter=",")
    for row in reader:
        lexicon_negative[row[0]] = int(row[1])

# Membuat fungsi untuk mempolarisasi sentimen berdasarkan kamus lexicon yang ada
def sentiment_analysis(text):
    score = 0
    for word in text:
        if (word in lexicon_positive):
            score = score + lexicon_positive[word]
    for word in text:
        if (word in lexicon_negative):
            score = score + lexicon_negative[word]
    polarity = ''
    if (score > 0 ):
        polarity = "Positif"
    elif (score < 0 ):
        polarity = "Negatif"
    else :
        polarity = "Netral"
    return score, polarity

[ ] result = df['stemming'].apply(sentiment_analysis)
result = list(zip(*result))
df["polarity_score"] = result[0]
df["polarity"] = result[1]
print(df["polarity"].value_counts())
df.shape

Negatif    1591
Netral      509
Positif     462
Name: polarity, dtype: int64
(2562, 10)

```

Gambar 3. Hasil Labelling Data

Modeling data berdasarkan algoritma yang telah ditentukan yaitu Multinomial Naïve Bayes dan Adaboost untuk menemukan hasil prediksi klasifikasi kelas sentiment berdasarkan kedua model algoritma tersebut. Pembagian data latih dan data uji pada tahap pemodelan dilakukan secara random dan didapatkan jumlah data latih sebanyak 2049 data Dan data jumlah data uji 513 data.

```
[ ] from sklearn.naive_bayes import MultinomialNB
    from sklearn.ensemble import AdaBoostClassifier
```

```
[ ] model = MultinomialNB()
    model.fit(tf_idf_train, y_train)
```

```
▾ MultinomialNB
MultinomialNB()
```

```
[ ] model_2 = AdaBoostClassifier()
    model_2.fit(tf_idf_train, y_train)
```

```
▾ AdaBoostClassifier
AdaBoostClassifier()
```

Gambar 4. Model Agoritma

Evaluasi model yang digunakan pada kedua metode ini berupa nilai confusion matrix dengan ukuran 3x3, yang berisi nilai akurasi, presisi, f1-score dan recall dari masing-masing kelas. Nilai tersebut didapatkan dari hasil pengolahan data train. Dalam menentukan nilai confusion matrix, penelitian ini menggunakan cross validation agar nilai yang dihasilkan maksimal. Rumus perhitungan tabel confusion matrix multidimensi adalah sebagai berikut:

1. Kelas negative

Aktual Data	Prediksi Data		
	Negatif	Netral	Positif
Negatif	TP	FN	FN
Netral	FP	TN	TN
Positif	FP	TN	TN

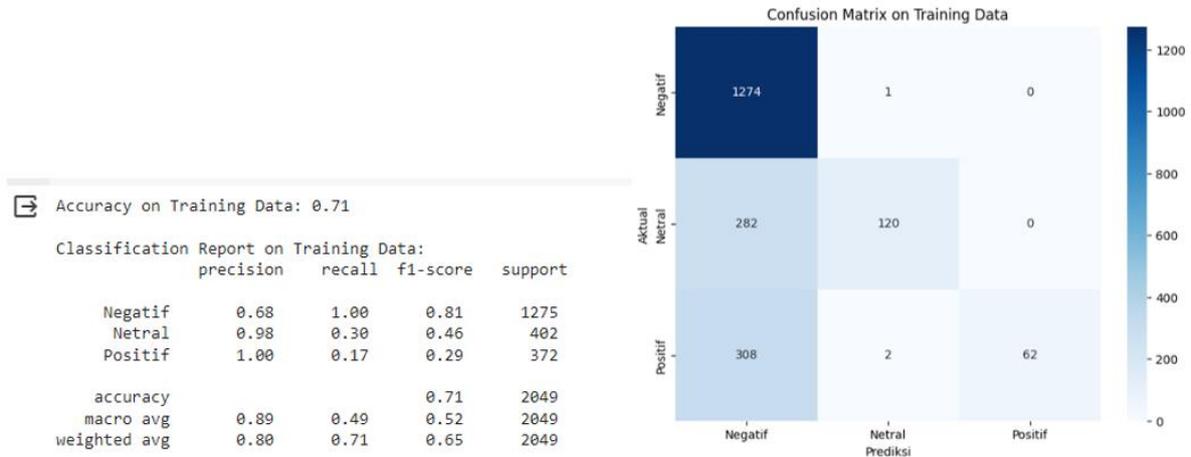
2. Kelas netral

Aktual Data	Prediksi Data		
	Negatif	Netral	Positif
Negatif	TP	FN	FN
Netral	FP	TN	TN
Positif	FP	TN	TN

3. Kelas positif

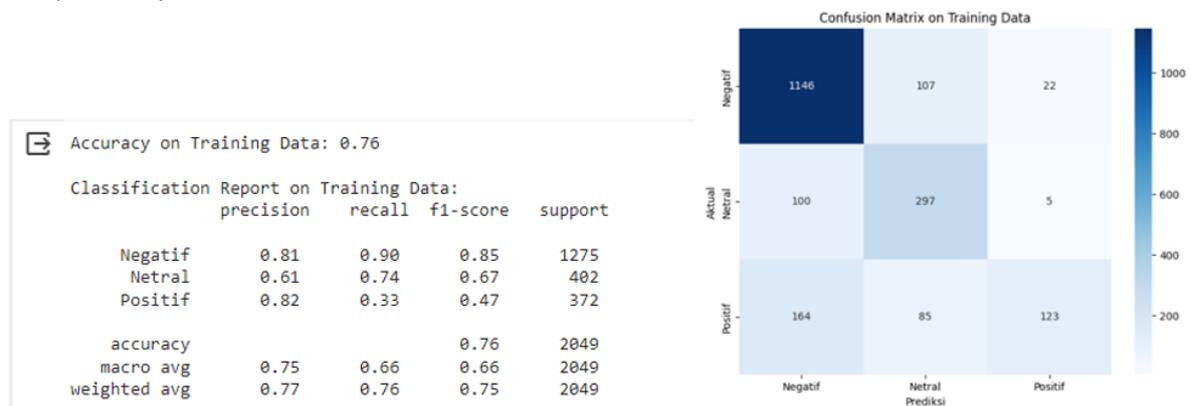
Aktual Data	Prediksi Data		
	Positif	Netral	Positif
Negatif	TN	TN	FP
Netral	TN	TN	FP
Positif	FN	FN	TP

Hasil evaluasi dari model algoritma multinomial naïve bayes didapatkan hasil nilai akurasi sebesar 0.71 atau 71%. nilai presisi untuk kelas negatif 68%, recall sebesar 100% dan f1-score sebesar 81%. Untuk kelas positif nilai presisinya sebesar 100%, recall sebesar 17% dan f1-score sebesar 29%. Dan untuk kelas netral nilai presisinya sebesar 98%, recall sebesar 30% dan f1-score sebesar 46%.



Gambar 5. Nilai Akurasi dan Confusion Matrix Multinomial Naïve Bayes

Hasil model algoritma Adaboost didapatkan nilai akurasi sebesar 76%, nilai presisi untuk kelas negatif 81%, recall sebesar 90% dan f1-score sebesar 85%. Untuk kelas positif nilai presisinya sebesar 82%, recall sebesar 33% dan f1-score sebesar 47%. Dan untuk kelas netral nilai presisinya sebesar 61%, recall sebesar 74% dan nilai f1-score sebesar 67%.



Gambar 6. Nilai Akurasi dan Confusion Matrix Adaboost

5. KESIMPULAN

Nilai akurasi, presisi, recall dan f1-score pada algoritma Adaboost lebih tinggi jika dibandingkan dengan algoritma Multinomial Naïve Bayes, yang mana algoritma Adaboost mendapatkan nilai akurasi sebesar 76% sedangkan algoritma Multinomial Naïve Bayes mendapatkan nilai akurasi sebesar 71%. Terbukti bahwa algoritma Adaboost bekerja lebih baik dalam mengolah data sentiment twitter terkait isu pinjaman online. Hasil klasifikasi sentiment dengan menggunakan kamus lexicon bahasa Indonesia didapatkan bahwa terdapat 462 data berlabel positif, 1.591 data berlabel negatif, dan terdapat 509 data berlabel netral.

Hal ini membuktikan bahwa keberadaan pinjaman online dari komentar masyarakat di media sosial twitter berdampak buruk bagi masyarakat.

REFERENSI

- [1] A. Z. Nihayah, I. Kahrismasuci, M. R. Chamami, and L. H. Rifqi, "Edukasi Keuangan Digital dalam Memanfaatkan Jasa Pinjaman Online," *Bubungan Tinggi J. Pengabd. Masy.*, vol. 5, no. 1, p. 231, 2023, doi: 10.20527/btjpm.v5i1.7325.
- [2] A. F. I. Muhammad Imam Ghozali, Wibowo Harry Sugiharto, "Analisis Sentimen Pinjaman Online Di Media Sosial Twitter Menggunakan Metode Naive Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 33, no. 1, pp. 1–12, 2022, doi: 10.30865/klik.v3i6.936.
- [3] M. Zaki Hariansyah, "Implementasi Metode Multinomial Naive Bayes pada Analisis Sentimen Terhadap Layanan Aplikasi Livin by Mandiri Implementation of Naive Bayes Multinomial Method on Sentiment Analysis of Livin by Mandiri Application Services," *Semin. Nas. Mhs. Fak. Teknol. Inf. Jakarta-Indonesia*, no. September, pp. 517–524, 2022, [Online]. Available: <https://senafti.budiluhur.ac.id/index.php>
- [4] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-Janua, no. December, pp. 391–394, 2017, doi: 10.1109/IALP.2017.8300625.
- [5] Y. Azhar, "Metode Lexicon-Learning Based Untuk Identifikasi Tweet Opini Berbahasa Indonesia," *J. Nas. Pendidik. Tek. Inform.*, vol. 6, no. 3, p. 237, 2018, doi: 10.23887/janapati.v6i3.11739.
- [6] A. Sentia, "Multinomial Naive Bayes Classifier Untuk Analisis Sentimen," no. December, pp. 0–8, 2023.
- [7] Zelin Gaa Ngilo and Nuryuliani Nuryuliani, "Analisis Sentimen Opini Pengguna Twitter Pada Aplikasi Bibit Menggunakan Multinomial Naive Bayes," *J. Tek. dan Sci.*, vol. 2, no. 1, pp. 08–15, 2023, doi: 10.56127/jts.v2i1.521.
- [8] F. Syah, H. Fajrin, A. N. Afif, M. R. Saeputra, D. Mirranty, and D. D. Saputra, "Analisa Sentimen Terhadap Twitter IndihomeCare Menggunakan Perbandingan Algoritma Smote, Support Vector Machine, AdaBoost dan Particle Swarm Optimization," *J. JTik (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 1, pp. 53–58, 2023, doi: 10.35870/jtik.v7i1.686.
- [9] A. Wahyu, N. Faizi, and K. Nugroho, "Penerapan Metode Adaptive Boosting Pada Analisis Sentimen Kenaikan BBM Pertamina," vol. 08, no. 2016, pp. 171–180, 2023.
- [10] A. P. Natasuwarna, "Tantangan Menghadapi Era Revolusi 4.0 - Big Data dan Data Mining," *Sindimas*, vol. 1, no. 1, pp. 23–27, 2019, [Online]. Available: <https://www.sisfotenika.stmikpontianak.ac.id/index.php/sindimas/article/view/530>
- [11] M. Munarsih and B. A. Ningsi, "Performance Comparison of Data Mining Classification Algorithms on Student Academic Achievement Prediction," *Indones. J. Artif. Intell. Data Min.*, vol. 6, no. 1, p. 29, 2023, doi: 10.24014/ijaidm.v6i1.21874.
- [12] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, "Text mining in big data analytics," *Big Data Cogn. Comput.*, vol. 4, no. 1, pp. 1–34, 2020, doi: 10.3390/bdcc4010001.
- [13] A. Sabrani, I. W. Gede Putu Wirarama Wedashwara, and F. Bimantoro, "Metode Multinomial Naive Bayes Untuk Klasifikasi Artikel Online Tentang Gempa Di Indonesia (Multinomial Naive Bayes Method for Classification of Online Article About Earthquake

- in Indonesia),” *Jtika*, vol. 2, no. 1, pp. 91–92, 2020, [Online]. Available: <http://jtika.if.unram.ac.id/index.php/JTIKA/>
- [14] F. Hadaina and U. Budiyanto, “Implementasi Metode Multinomial Naïve Bayes Untuk Sentiment Analysis Terhadap Data Ulasan Produk Colearn Pada Google Play Store Implementation Of Multinomial Naive Bayes Method For Sentiment Analysis Of Colearn Product Review Data On Google Play Store,” *Semin. Nas. Mhs. Fak. Teknol. Inf. Jakarta-Indonesia*, no. September, pp. 660–666, 2022, [Online]. Available: <https://senafti.budiluhur.ac.id/index.php>
- [15] R. Rabbani, I. Wahidah, and I. H. Santoso, “Klasifikasi Data Deteksi Jatuh Menggunakan Machine Learning dengan Algoritma Adaptive Boosting (AdaBoost),” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 5053–5063, 2021.