

PERBANDINGAN ALGORITMA RANDOM FOREST DAN XGBOOST UNTUK ANALISIS SENTIMEN PUBLIK TERHADAP RUMAH MAKAN PAYAKUMBUAH

Hayyun Lestia Wati¹, Moch TBG Ilham Ramdhan Wiradinata², Neni Anggraeni³, Siti Kolbiah⁴,
Ujang Hendar⁵, Nova Agustina⁶
Departemen Teknik Informatika^{1, 2, 3, 4, 5, 6}
Universitas Teknologi Bandung^{1, 2, 3, 4, 5, 6}
lestihayyun@gmail.com¹, ilham.wiradinata93@gmail.com², anggraenin378@gmail.com³, sitikolbiah5@gmail.com⁴,
ujanghendar738@gmail.com⁵, nova@utb-univ.ac.id⁶

Abstrak

Penelitian ini bertujuan membandingkan efektivitas dua algoritma *machine learning*, yaitu Random Forest dan XGBoost, dalam analisis sentimen terhadap ulasan publik mengenai Rumah Makan Payakumbuh. *Dataset* yang digunakan terdiri dari 1.186 ulasan yang dikumpulkan dari berbagai platform media sosial. Sebelum proses klasifikasi dilakukan, data terlebih dahulu melalui tahapan *preprocessing* dan transformasi teks menggunakan metode TF-IDF untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma. Eksperimen dilakukan dengan menerapkan kedua algoritma untuk mengklasifikasikan sentimen ulasan menjadi kategori positif, negatif, atau netral. Hasil menunjukkan bahwa XGBoost memberikan performa yang lebih baik dibandingkan Random Forest. XGBoost mencapai akurasi 68%, *precision* 66%, *recall* 68%, dan *F1-score* 66%. Sementara itu, Random Forest menghasilkan akurasi 63%, *precision* 61%, *recall* 63%, dan *F1-score* 61%. Meskipun keduanya mengalami kesulitan dalam mengklasifikasikan sentimen netral, XGBoost menunjukkan ketahanan yang lebih baik terhadap ketidakseimbangan kelas. Selain klasifikasi, penelitian ini juga melakukan analisis visual menggunakan *word cloud* untuk mengidentifikasi kata-kata yang paling sering muncul dalam ulasan. Hasilnya menunjukkan bahwa aspek rasa makanan dan kualitas pelayanan merupakan topik yang paling banyak dibahas oleh pelanggan. Temuan ini memberikan kontribusi penting dalam pengembangan sistem analisis sentimen yang lebih akurat dan adaptif untuk sektor kuliner. Dengan memahami pola umpan balik pelanggan secara lebih mendalam, pelaku usaha dapat meningkatkan kualitas layanan dan merespons kebutuhan konsumen secara lebih efektif. Penelitian ini juga membuka peluang untuk pengembangan model yang lebih kompleks di masa depan, khususnya dalam menangani data ulasan yang tidak seimbang secara kelas.

Kata kunci : Analisis Sentimen, Random Forest, XGBoost, Rumah Makan Payakumbuh, *machine learning*

Abstract

This study aims to compare the effectiveness of two machine learning algorithms—Random Forest and XGBoost—in sentiment analysis of public reviews about Payakumbuh Restaurant. The dataset comprises 1,186 reviews collected from various social media platforms. Prior to classification, the data underwent preprocessing and text transformation using TF-IDF to convert textual data into numerical representations processable by the algorithms. Experiments applied both algorithms to classify reviews into positive, negative, or neutral sentiments. Results demonstrate XGBoost's superior performance over Random Forest, achieving 68% accuracy, 66% precision, 68% recall, and 66% F1-score, compared to Random Forest's 63% accuracy, 61% precision, 63% recall, and 61% F1-score. Although both algorithms struggled with neutral sentiment classification, XGBoost exhibited better resilience to class imbalance. Beyond classification, this research also conducted visual analysis using word clouds to identify frequently occurring review terms. Findings reveal that food flavor and service quality were the most discussed topics among customers. These results contribute significantly to developing more accurate and adaptive sentiment analysis systems for the culinary sector. By deeply understanding customer feedback patterns, businesses can enhance service quality and respond to consumer needs more effectively. This study also opens opportunities for developing more complex future models, particularly in handling class-imbalanced review data.

Keywords : Sentiment Analysis, Random Forest, XGBoost, Payakumbuh Restaurant, machine learning

I. PENDAHULUAN

Ulasan dan opini masyarakat terhadap suatu produk atau layanan menjadi faktor krusial, terutama dalam industri kuliner. Rumah makan, yang berfokus pada kepuasan pelanggan, sangat bergantung pada masukan serta persepsi publik guna meningkatkan kualitas layanan dan produk mereka. Seiring dengan meningkatnya penggunaan media sosial dan platform ulasan daring, analisis sentimen terhadap rumah makan menjadi semakin relevan dalam memahami preferensi pelanggan serta tren pasar. Salah satu rumah makan yang menarik perhatian publik adalah Rumah Makan Payakumbuh, yang dimiliki oleh Arief Muhammad, seorang *content creator* dan selebritas media sosial yang dikenal sebagai "duta nasi padang" di Instagram. Rumah makan ini mengusung konsep makanan khas Minang dengan sentuhan modern. Arief Muhammad resmi membuka *outlet* pertamanya pada 14 Juli 2022 di kawasan Ciater Raya, Serpong, Tangerang Selatan, setelah menerima gelar kehormatan dari Wakil Gubernur Sumatera Barat, Audy Joinaldy. Dengan strategi pemasaran yang kuat melalui media sosial, Rumah Makan Payakumbuh berhasil menarik banyak pelanggan serta memperoleh berbagai ulasan dari konsumen. Oleh karena itu, analisis sentimen terhadap rumah makan ini menjadi penting guna memahami bagaimana opini publik dapat memengaruhi popularitas dan strategi bisnisnya.

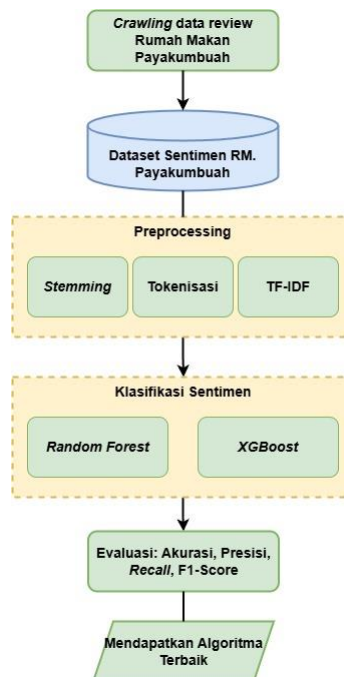
Analisis sentimen merupakan cabang dari pemrosesan bahasa alami (NLP) yang bertujuan untuk mengkategorikan opini ke dalam beberapa klasifikasi, seperti positif, negatif, atau netral [1], [2]. Penelitian ini bertujuan untuk melakukan analisis sentimen publik terhadap Rumah Makan Payakumbuh dengan menerapkan dua algoritma pembelajaran mesin

yang populer, yaitu Random Forest [3], [4] dan *Extreme Gradient Boosting* (XGBoost) [5]. Kedua metode ini banyak digunakan karena kemampuan metode dalam menangani data berdimensi tinggi, mengenali pola non-linear, serta memberikan hasil yang kompetitif dalam berbagai jenis *dataset*. XGBoost merupakan salah satu metode *boosting* yang unggul karena mengusung optimasi berbasis regularisasi untuk meningkatkan akurasi dan mengurangi risiko *overfitting* [6]. Algoritma ini dikenal memiliki efisiensi tinggi serta akurasi yang sangat baik dalam menangani data dalam jumlah besar, sekaligus mampu mengatasi permasalahan bias dan varians lebih efektif dibandingkan metode lainnya.

Meskipun kedua algoritma memiliki keunggulan masing-masing, perbandingan kinerjanya dalam konteks analisis sentimen terhadap Rumah Makan Payakumbuh masih jarang diteliti. Oleh karena itu, penelitian ini berfokus pada evaluasi serta perbandingan performa Random Forest dan XGBoost dalam menganalisis sentimen publik berdasarkan data ulasan dari berbagai platform daring. Hasil penelitian ini diharapkan dapat memberikan wawasan terkait algoritma yang lebih optimal dalam analisis sentimen publik serta memberikan rekomendasi bagi pemilik usaha kuliner dalam memahami opini pelanggan dan meningkatkan kualitas layanan mereka.

II. TINJAUAN PUSTAKA

Metode penelitian bertujuan untuk mengulas beberapa topik penting dalam Ilmu Komputer dan Informatika, dengan fokus pada perbandingan dua algoritma populer, Random Forest dan XGBoost, dalam konteks analisis sentimen terhadap publik yang mengunjungi rumah makan di Payakumbuh. Penelitian analisis sentimen telah banyak diterapkan dalam berbagai bidang, termasuk kuliner, dengan menggunakan berbagai algoritma *machine learning* [7], [8]. Beberapa studi sebelumnya menunjukkan bahwa algoritma seperti Random Forest dan XGBoost sering digunakan karena kemampuannya menangani data berdimensi tinggi dan klasifikasi multikelas. Random Forest dikenal stabil terhadap *noise* [9], sementara XGBoost menawarkan optimasi dan performa yang tinggi melalui teknik *boosting* [10]. Namun, hingga saat ini, masih terdapat keterbatasan dalam studi yang secara langsung membandingkan kedua algoritma ini dalam konteks analisis sentimen terhadap ulasan rumah makan lokal, khususnya yang berasal dari Indonesia. Sehingga, membandingkan efektivitas algoritma Random Forest dan XGBoost dalam menganalisis sentimen publik terhadap Rumah Makan Payakumbuh menjadi gap penelitian ini. Kerangka pemikiran pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Pemikiran

1. Analisis Sentimen

Analisis sentimen adalah proses yang digunakan untuk menganalisis teks atau opini yang diungkapkan dalam bentuk tulisan dan mengklasifikasikannya ke dalam kategori positif, negatif, atau netral [11], [12]. Teknologi ini telah banyak digunakan di berbagai bidang seperti pemasaran, layanan pelanggan, dan penelitian sosial untuk memahami pandangan dan perasaan publik. Analisis sentimen dapat menggunakan teknik pengolahan bahasa alami (*Natural Language Processing/NLP*) untuk mengidentifikasi dan mengekstrak informasi dari teks [13], [14], [15]. Dalam hal ini, sentimen yang diekstraksi dapat mencerminkan opini pengguna tentang produk atau

layanan tertentu, seperti rumah makan. Analisis sentimen menjadi alat penting dalam memonitor kepuasan pelanggan dan meningkatkan kualitas layanan.

2. Random Forest

Random Forest adalah algoritma pembelajaran berbasis pohon keputusan (*decision tree*) yang digunakan untuk klasifikasi dan regresi [16], [17]. Algoritma ini bekerja dengan membangun banyak pohon keputusan dari *subset* acak dari data pelatihan dan kemudian menggabungkan hasil dari masing-masing pohon untuk meningkatkan akurasi dan mengurangi *overfitting*. Kelebihan lainnya adalah kemampuan algoritma ini untuk menangani data dengan variabel input yang banyak dan fitur yang beragam. Dalam konteks analisis sentimen, Random Forest dapat digunakan untuk mengklasifikasikan teks atau opini berdasarkan fitur-fitur seperti frekuensi kata, kata kunci, atau polaritas kalimat. Algoritma ini memiliki performa yang baik dalam mengklasifikasikan sentimen positif atau negatif yang terdapat dalam ulasan atau komentar publik tentang rumah makan.

3. XGBoost

Extreme Gradient Boosting (XGBoost) adalah algoritma pembelajaran mesin berbasis pohon keputusan yang lebih canggih dan efisien dibandingkan dengan Random Forest. XGBoost menggunakan teknik *gradient boosting*, yang membangun model secara berurutan dengan menambahkan pohon keputusan baru untuk mengurangi kesalahan dari pohon sebelumnya [5], [6]. Keunggulan utama XGBoost adalah kemampuannya untuk menangani data yang besar dan kompleks dengan kecepatan yang tinggi serta menghasilkan model yang lebih akurat. XGBoost juga memiliki teknik *regularization* yang dapat membantu menghindari *overfitting*, sehingga sering digunakan dalam kompetisi data science dan aplikasi dunia nyata. Dalam analisis sentimen, XGBoost telah terbukti memberikan hasil yang lebih akurat dibandingkan dengan algoritma pembelajaran mesin lainnya karena kemampuannya dalam memodelkan hubungan non-linear yang lebih kompleks. XGBoost bekerja dengan cara mengoptimalkan fungsi kerugian dengan menggunakan gradien untuk memperbaiki kesalahan prediksi. Algoritma ini sering digunakan dalam klasifikasi sentimen untuk mendeteksi dan mengklasifikasikan sentimen dari berbagai jenis teks, seperti ulasan restoran atau rumah makan.

4. Perbandingan Algoritma Random Forest dan XGBoost

Banyak penelitian yang telah membandingkan performa Random Forest dan XGBoost dalam berbagai aplikasi analisis data, termasuk analisis sentimen. Berdasarkan beberapa penelitian [10], [18], XGBoost sering kali unggul dalam hal akurasi prediksi dibandingkan dengan Random Forest, terutama pada *dataset* yang lebih besar dan lebih kompleks. Hasil tersebut membuktikan XGBoost mampu mengoptimalkan model secara berurutan dan menggunakan teknik *regularisasi*. Namun, Random Forest tetap menjadi pilihan yang baik untuk *dataset* yang lebih kecil atau ketika kecepatan pemrosesan dan interpretabilitas model menjadi prioritas. Selain itu, Random Forest lebih mudah digunakan dalam situasi di mana pemilihan fitur secara otomatis dan efisiensi waktu tidak terlalu kritis. Dalam analisis sentimen terhadap rumah makan, baik Random Forest maupun XGBoost dapat digunakan dengan baik, tetapi pilihan antara keduanya bergantung pada ukuran *dataset*, kompleksitas data, dan kebutuhan akurasi. Sebagai contoh, jika *dataset* terdiri dari banyak ulasan dengan nuansa sentimen yang lebih kompleks, XGBoost mungkin lebih baik, sementara Random Forest bisa cukup efektif untuk *dataset* yang lebih sederhana.

III. METODE PENELITIAN

Pada penelitian ini, dilakukan perbandingan antara dua algoritma populer dalam analisis sentimen, yaitu Random Forest dan XGBoost, yang diterapkan pada *dataset* ulasan publik terhadap Rumah Makan Payakumbuh. Tujuan utama adalah untuk mengevaluasi kinerja kedua algoritma ini dalam mengklasifikasikan sentimen publik yang berupa positif, negatif, atau netral berdasarkan teks ulasan.

1. Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 1186 ulasan publik yang diberikan oleh pengguna berbagai platform media sosial. Setiap ulasan memiliki label sentimen yang sudah ditentukan, yaitu positif, negatif, dan netral. Kolom yang tersedia dalam *dataset* adalah *username*, *text*, sumber, dan label. *Dataset* ini diproses menggunakan teknik pemrosesan teks, seperti penghilangan tanda baca, perubahan teks menjadi huruf kecil, dan penghapusan kata yang tidak relevan.

| | A | B | C | D | E |
|----|----|------------------|--|-----------|---------|
| 1 | no | username | text | sumber | label |
| 2 | 1 | fajarpandun | Udh punya lisensi dari IKM blom? *Ikatan Keluarga Minang | Instagram | Netral |
| 3 | 2 | anne.w.d | d diemin aja g ngmg apa2. Sampe lama n sampe tanya sndri klo ternyata nasinya harus by order. Lha gtu knp g bilang | Instagram | Negatif |
| 4 | 3 | salsahehe | Yg cabang jogja buka kapan ya minn? Dah laper ini | Instagram | Netral |
| 5 | 4 | harunakbar90 | lebih penting Lisensi Dr IKM | Instagram | Negatif |
| 6 | 5 | bluetifull_21 | ar90 Sepenting apasih? Bukannya armuh udah diangkat jadi duta nasi padang. Spam bangetdeh malu"n warga asli m | Instagram | Negatif |
| 7 | 6 | khnsaydh | aaa mauu coba | Instagram | Positif |
| 8 | 7 | salwasyaarani | Duhhh harus coba ini | Instagram | Positif |
| 9 | 8 | diitaatiningsih | Boleh nggak ya bwa minum sendiri | Instagram | Netral |
| 10 | 9 | putri.skrw | Harus nyobain si ini | Instagram | Positif |
| 11 | 10 | intanshanty | Minim 100 yaa? | Instagram | Netral |
| 12 | 11 | latifshulaiman | itu ayam pop nya kalau di makan nyanyi gak? | Instagram | Negatif |
| 13 | 12 | dyahellypalupi | Telur barendonya gak menarik... | Instagram | Negatif |
| 14 | 13 | noenk_ariadi | Mau mampir ngga jadi, krn ngga ada sate padangnya... | Instagram | Negatif |
| 15 | 14 | aratiaryatyas | ngilerrrr | Instagram | Positif |
| 16 | 15 | fachmisetyawan | Standar makan 1 orang 30-40 ya | Instagram | Netral |
| 17 | 16 | hartanto83 | Tiap hati lewat tp gak fomo lbh ke kaum mendang mending sih gue kalau makan nasi padang di situ | Instagram | Negatif |
| 18 | 17 | sunu_raharja | Berdua yo 1 jt... masih aman... | Instagram | Netral |
| 19 | 18 | nursyawalh | 11/100 ini mah. Sangat ga worth it | Instagram | Negatif |
| 20 | 19 | nabaandre | Pantes pas makan di sana nasi nya ketemu kuah gulali jd makin mantab | Instagram | Positif |
| 21 | 20 | gregoriusruben | Terbaik ini | Instagram | Positif |
| 22 | 21 | bimacommerce | Jadi pengen buat nyobain | Instagram | Positif |
| 23 | 22 | aliefdky | Enak, tapi overated | Instagram | Positif |
| 24 | 23 | iwa_mylife | B aja. | Instagram | Negatif |
| 25 | 24 | siomayaan.id | Viral sesaat aja...hitungan bulan udah sepi. | Instagram | Negatif |
| 26 | 25 | daviddrmdhn | tetep duta minang gondolayu | Instagram | Positif |
| 27 | 26 | astribangjoe | Kidul omah gur mlaku iso wkwkwkw | Instagram | Positif |
| 28 | 27 | suryowsaputro | astribangjoe bar mangan lgsg boncoss | Instagram | Negatif |
| 29 | 28 | totoksubiyantoro | Tetep lurr... kanggo aku kaum mendang mending RM padang giwangan juarane | Instagram | Negatif |
| 30 | 29 | bielmuh | Terdebeess Masyugg | Instagram | Positif |
| 31 | 30 | budie_ariief | duwh gaji umr minggir dulu , pilih nasi padang paket 12rebu pinggir jalan wae | Instagram | Negatif |

Gambar 2. Dataset Rumah Makan Payakumbuh

2. Preprocessing dan Pembagian Data

Dalam implementasi analisis sentimen terhadap Rumah Makan Payakumbuh, tahap *preprocessing* data memegang peranan krusial untuk memastikan kualitas input yang akan diproses oleh model *machine learning*. Proses ini diawali dengan pembuatan fungsi *preprocess_text* yang melakukan dua operasi utama: merubah seluruh teks menjadi huruf kecil (*lowercase*) dan menghapus karakter-karakter non-alfabetik menggunakan *regular expression*. Konversi ke huruf kecil bertujuan untuk membuat standar teks, sementara penghapusan karakter non-alfabetik membantu mengurangi noise dan meningkatkan fokus analisis pada konten tekstual yang relevan.

```
# Preprocessing data
def preprocess_text(text):
    text = text.lower()
    text = re.sub(r'^a-zA-Z\s', '', text)
    return text

df['text'] = df['text'].astype(str).apply(preprocess_text)
```

Gambar 3. Preprocessing Data

Setelah mendefinisikan fungsi *preprocessing*, langkah selanjutnya adalah penerapan fungsi tersebut pada *dataset*. Kolom *'text'* yang berisi ulasan pengguna dikonversi terlebih dahulu ke tipe data string menggunakan method *astype(str)*, kemudian fungsi *preprocess_text* diterapkan ke setiap baris data menggunakan method *apply()*. Proses ini memastikan konsistensi format data dan menghilangkan variasi yang tidak diperlukan dalam analisis sentimen.

```
# Split data
X_train, X_test, y_train, y_test = train_test_split(df['text'], df['label'], test_size=0.2, random_state=42)
```

Gambar 4. Split Data Train dan Test

Tahap terakhir dalam persiapan data adalah pembagian *dataset* menjadi data training dan testing menggunakan fungsi *train_test_split* dari *library scikit-learn*. *Dataset* dibagi dengan proporsi 80:20, di mana 80% data digunakan untuk melatih model (training) dan 20% sisanya digunakan untuk evaluasi (testing). Parameter *random_state=42* ditetapkan untuk menjamin *reproducibility*, memastikan hasil pembagian data konsisten setiap kali kode dijalankan. Kolom *'text'* yang telah diproses digunakan sebagai fitur (X), sementara kolom *'label'* yang berisi klasifikasi sentimen digunakan sebagai target (y). Pembagian data ini fundamental untuk mengevaluasi performa model secara objektif dan menghindari *overfitting*.

3. Pengaturan Model dan Hyperparameter

Pada penelitian ini, dua algoritma *machine learning* digunakan untuk eksperimen, yaitu Random Forest dan XGBoost. Sebelum melatih model, data teks diolah menggunakan *TfidfVectorizer* untuk mengubah data teks mentah menjadi representasi numerik berupa matriks *Term Frequency-Inverse Document Frequency* (TF-IDF),

yang bertujuan untuk menilai pentingnya suatu kata dalam dokumen tertentu relatif terhadap semua dokumen dalam dataset.

IV. HASIL DAN ANALISIS

1. Random Forest

Model Random Forest dikonfigurasi dengan 100 pohon keputusan ($n_estimators=100$) dan parameter *default* lainnya, serta menggunakan $random_state=42$ untuk memastikan hasil yang konsisten. Proses pelatihan dilakukan menggunakan data latih yang telah ditransformasikan menjadi matriks TF-IDF, dan model menghasilkan prediksi pada data uji dengan metode prediksi.

```
# Feature extraction
vectorizer = TfidfVectorizer()
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)

# Train Random Forest model
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train_tfidf, y_train)

# Predictions
y_pred = model.predict(X_test_tfidf)
```

Gambar 5. Model Random Forest

2. XGBoost

Model XGBoost dikonfigurasi dengan 100 estimasi ($n_estimators=100$), laju pembelajaran sebesar 0.1 ($learning_rate=0.1$), dan metrik evaluasi *multiclass log loss*. Untuk menyesuaikan versi terbaru XGBoost, pengaturan $use_label_encoder=False$ digunakan. Model ini juga dilatih dengan data latih yang telah ditransformasikan ke bentuk TF-IDF, kemudian melakukan prediksi pada data uji.

```
# Feature extraction
vectorizer = TfidfVectorizer()
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)

# Train XGBoost model
model = xgb.XGBClassifier(n_estimators=100, learning_rate=0.1, random_state=42, use_label_encoder=False, eval_metric='mlogloss')
model.fit(X_train_tfidf, y_train)

# Predictions
y_pred = model.predict(X_test_tfidf)
```

Gambar 6 Model XGBoost

3. Hasil Klasifikasi

Berikut adalah hasil evaluasi untuk kedua algoritma berdasarkan beberapa metrik utama, yaitu akurasi, *precision*, *recall*, *F1-score*, dan *support*.

a. Random Forest

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Negatif | 0.63 | 0.70 | 0.67 | 94 |
| Netral | 0.39 | 0.25 | 0.30 | 48 |
| Positif | 0.69 | 0.74 | 0.71 | 96 |
| accuracy | | | 0.63 | 238 |
| macro avg | 0.57 | 0.56 | 0.56 | 238 |
| weighted avg | 0.61 | 0.63 | 0.61 | 238 |

Gambar 7. Hasil Klasifikasi Random Forest

Random Forest menunjukkan performa yang bervariasi dalam mengklasifikasikan sentimen. Untuk sentimen negatif, model mencapai *precision* 63% yang menunjukkan akurasi prediksi negatif, dengan *recall* 70% yang mengindikasikan kemampuan model dalam mengidentifikasi data negatif, menghasilkan *f1-score* 67% dari 94 sampel. Pada klasifikasi sentimen netral, model menunjukkan performa yang lebih rendah dengan *precision* 39% dan *recall* 25%, menghasilkan *f1-score* 30% dari 48 sampel, yang merupakan performa terendah di antara semua kelas. Untuk sentimen positif, model mencapai performa terbaik dengan *precision* 69% dan *recall* 74%, menghasilkan *f1-score* 0.71 dari 96 sampel.

b. XGBoost

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Negatif | 0.67 | 0.79 | 0.73 | 94 |
| Netral | 0.50 | 0.25 | 0.33 | 48 |
| Positif | 0.72 | 0.78 | 0.75 | 96 |
| accuracy | | | 0.68 | 238 |
| macro avg | 0.63 | 0.61 | 0.60 | 238 |
| weighted avg | 0.66 | 0.68 | 0.66 | 238 |

Gambar 8. Hasil Klasifikasi XGBoost

XGBoost memperlihatkan peningkatan performa dibandingkan Random Forest dalam mengklasifikasikan sentimen. Untuk sentimen negatif, model mencapai *precision* 67% dengan *recall* yang lebih tinggi yaitu 79%, menghasilkan *f1-score* 73% dari 94 sampel. Pada klasifikasi sentimen netral, meskipun masih menjadi tantangan, model menunjukkan sedikit peningkatan dengan *precision* 50% meski *recall* tetap di 25%, menghasilkan *f1-score* 33% dari 48 sampel. Sentimen positif menunjukkan performa terbaik dengan *precision* 72% dan *recall* 78%, menghasilkan *f1-score* 75% dari 96 sampel.

c. Tabel Hasil Akurasi Algoritma Random Forest dan XGBoost

Berdasarkan hasil klasifikasi yang telah dianalisis sebelumnya, berikut tabel hasil klasifikasi untuk algoritma Random Forest dan XGBoost:

TABEL I
 HASIL AKURASI RANDOM FOREST DAN XGBOOST

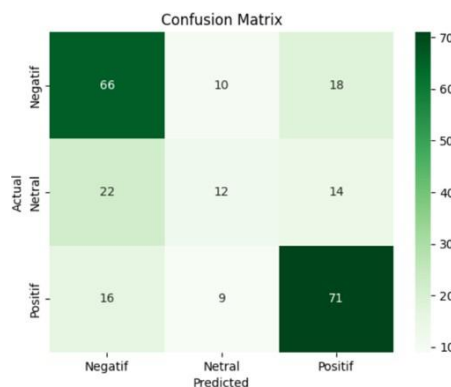
| Metric | Random Forest | XGBoost |
|------------------|---------------|---------|
| Akurasi | 63% | 68% |
| <i>Precision</i> | 61% | 66% |
| <i>Recall</i> | 63% | 68% |
| <i>F1-Score</i> | 61% | 66% |

Kami menggunakan nilai *weighted average* untuk *precision*, *recall*, dan *F1-Score* karena nilai ini memberikan gambaran performa keseluruhan model dengan mempertimbangkan jumlah sampel di setiap kelas. Nilai-nilai ini menunjukkan bahwa XGBoost secara konsisten memberikan performa yang lebih baik dibandingkan Random Forest di semua metrik evaluasi.

4. Confusion Matrix

Evaluasi performa model Random Forest dan XGBoost dilakukan menggunakan *confusion matrix* untuk mengukur akurasi prediksi pada tiga kelas: Negatif, Netral, dan Positif.

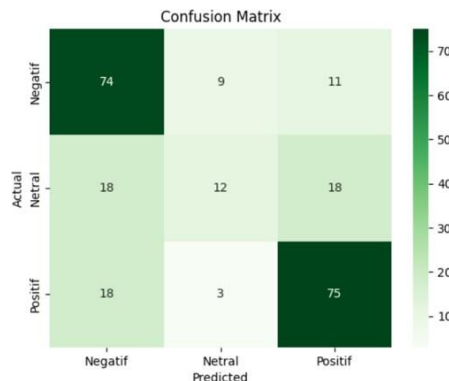
a. Random Forest



Gambar 9. Random Forest

Model Random Forest berhasil memprediksi 66 sampel dengan benar untuk kelas Negatif, 12 sampel untuk kelas Netral, dan 71 sampel untuk kelas Positif. Namun, model ini menunjukkan tingkat kesalahan yang relatif tinggi pada kelas Netral, dengan 22 sampel salah diklasifikasikan sebagai Negatif dan 14 sampel salah diklasifikasikan sebagai Positif.

b. XGBoost



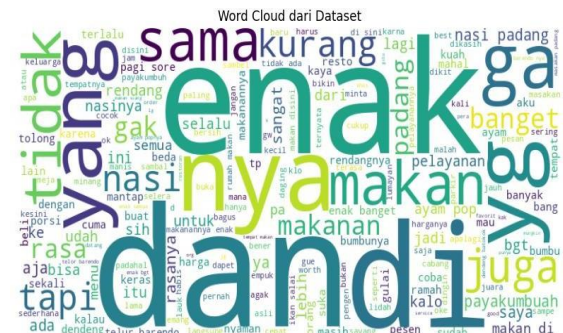
Gambar 10. XGBoost

5. Wordcloud

Pada penelitian ini, dilakukan analisis teks terhadap ulasan pelanggan menggunakan dua model *machine learning*, yaitu Random Forest dan XGBoost, untuk mengidentifikasi kata-kata yang sering muncul dalam ulasan. Hasil visualisasi dalam bentuk *word cloud* menunjukkan bahwa ulasan pelanggan didominasi oleh kata-kata yang berkaitan dengan makanan dan pelayanan, seperti "enak," "makan," "nasi," dan "rendang." Hal ini mengindikasikan bahwa aspek rasa makanan dan kualitas pelayanan merupakan faktor utama yang diperhatikan oleh pelanggan dalam memberikan ulasan. Selain itu, kata-kata seperti "kurang" dan "tidak" juga muncul dalam *word cloud*, yang menunjukkan adanya ulasan dengan sentimen negatif. Secara keseluruhan, hasil ini memberikan wawasan tentang faktor-faktor yang paling memengaruhi kepuasan pelanggan dan dapat digunakan sebagai dasar untuk meningkatkan kualitas layanan dan produk. *Word cloud* yang dihasilkan pada penelitian ini dapat dilihat pada Gambar 11 (Random Forest) dan 12 (XGBoost)



Gambar 11. Random Forest



Gambar 12. XGBoost

6. Analisis Perbandingan

Berdasarkan hasil eksperimen, XGBoost menunjukkan performa yang lebih unggul dibandingkan Random Forest dalam analisis sentimen ulasan Rumah Makan Payakumbuh. XGBoost mencapai akurasi keseluruhan 68%, sedangkan Random Forest mencapai 63%. Dalam klasifikasi sentimen negatif, XGBoost mencatatkan *F1-score* 0.73 (*precision* 67%, *recall* 79%), lebih tinggi dibanding Random Forest dengan *F1-score* 0.67 (*precision* 63%, *recall* 70%). Untuk sentimen positif, XGBoost kembali unggul dengan *F1-score* 0.75 dibanding Random Forest dengan 0.71.

Kedua algoritma menghadapi tantangan dalam mengklasifikasikan sentimen netral, dengan XGBoost mencapai *F1-score* 0.33 dan Random Forest 0.30. Analisis *confusion matrix* menunjukkan XGBoost lebih akurat dalam klasifikasi, dengan 74 sampel negatif, 12 netral, dan 75 positif terklasifikasi dengan benar, dibandingkan Random Forest yang mengklasifikasikan 66 negatif, 12 netral, dan 71 positif dengan benar.

Meskipun Random Forest memiliki keunggulan dalam waktu komputasi karena kemampuan pemrosesan paralel, XGBoost memberikan hasil yang lebih optimal dalam klasifikasi berkat mekanisme *gradient boosting* dan regularisasi yang lebih efektif. Hasil ini menunjukkan bahwa XGBoost merupakan pilihan yang lebih baik untuk sistem analisis sentimen yang memprioritaskan akurasi, sementara Random Forest tetap menjadi alternatif yang layak ketika efisiensi komputasi menjadi pertimbangan utama.

V. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa algoritma XGBoost memiliki performa yang lebih unggul dibandingkan dengan Random Forest dalam analisis sentimen publik terhadap Rumah Makan Payakumbuh. XGBoost mencapai akurasi 68%, *precision* 66%, *recall* 68%, dan *F1-score* 66%, sedangkan Random Forest memperoleh akurasi 63%,

precision 61%, *recall* 63%, dan *F1-score* 61%. Kedua algoritma menghadapi tantangan dalam mengklasifikasikan sentimen netral, namun XGBoost menunjukkan ketahanan yang lebih baik dalam menangani ketidakseimbangan kelas. Dari total 1186 ulasan yang dianalisis, mayoritas ulasan berfokus pada kualitas makanan dan pelayanan, dengan kata-kata dominan seperti “enak”, “makan” “nasi”, dan “rendang” yang muncul dalam *word cloud*. Implementasi *preprocessing* data menggunakan TF-IDF terbukti efektif dalam meningkatkan kualitas data sebelum analisis, serta pembagian *dataset* 80:20 untuk *training* dan *testing* memberikan hasil yang optimal.

REFERENSI

- [1] K. L. Tan, C. P. Lee, and K. M. Lim, “A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research,” *Applied Sciences (Switzerland)*, vol. 13, no. 7, 2023, doi: 10.3390/app13074550.
- [2] A. Athar, S. Ali, M. M. Sheeraz, S. Bhattacharjee, and H.-C. Kim, “Sentimental Analysis of Movie Reviews using Soft Voting Ensemble-based Machine Learning,” no. March, pp. 01–05, 2022, doi: 10.1109/snams53716.2021.9732159.
- [3] Bahwari, “Sentiment Analysis Using Random Forest Algorithm - Online Social Media Based,” *Journal Of Information Technology AND ITS UTILIZATION*, vol. 2, no. 2, pp. 29–33, 2019, [Online]. Available: https://www.researchgate.net/publication/338548518_SENTIMENT_ANALYSIS_USING_RANDOM_FOREST_ALGORITHM_ONLINE_SOCIAL_MEDIA_BASED
- [4] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, “Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB,” *Jurnal Informatika Upgris*, vol. 7, no. 1, pp. 36–40, 2021, doi: 10.26877/jiu.v7i1.7099.
- [5] suwarno and R. Kusnadi, “Analisis Perbandingan SVM, XGBoost dan Neural Network pada Klasifikasi,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informatika)*, vol. 5, no. 5, pp. 896–903, Oct. 2021, doi: 10.29207/resti.v5i5.3506.
- [6] X. Y. Liew, N. Hameed, and J. Clos, “An investigation of XGBoost-based algorithm for breast cancer classification,” *Machine Learning with Applications*, vol. 6, no. September, p. 100154, 2021, doi: 10.1016/j.mlwa.2021.100154.
- [7] M. F. Firdaus, D. E. Ratnawati, and N. Y. Setiawan, “Analisis Sentimen Berbasis Aspek Pada Ulasan Pelanggan Restoran Menggunakan Algoritma Support Vector Machine (Studi Kasus: Depot Bamara),” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1265–1272, Dec. 2024, doi: 10.25126/jtiik.2024117564.
- [8] D. F. Salsabillah, D. E. Ratnawati, and N. Y. Setiawan, “Analisis Sentimen Ulasan Rumah Makan Menggunakan Perbandingan Algoritma Support Vector Machine dengan Naive Bayes (Studi Kasus: Ayam Goreng Nelongso Cabang Singosari, Malang),” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 107–116, Feb. 2024, doi: 10.25126/jtiik.2024117584.
- [9] M. A. A. Far, R. Andreswari, and D. Pramesti, “Sentiment Analysis on Youtube Social Media Using Decision Tree and Random Forest Algorithm: A Case Study,” in *2020 International Conference on Data Science and Its Applications (ICoDSA)*, IEEE, Aug. 2020, pp. 1–7. doi: 10.1109/ICoDSA50139.2020.9213078.
- [10] D. Wijayanto and Bambang Pulu Hartato, “Analisis Perbandingan Performa Algoritma XGBoost dan LightGBM pada Klasifikasi Kanker Payudara,” *The Indonesian Journal of Computer Science*, vol. 13, no. 2, pp. 2337–3520, Apr. 2024, doi: 10.33022/ijcs.v13i2.3901.
- [11] C. J. Varshney, A. Sharma, and D. P. Yadav, “Sentiment analysis using ensemble classification technique,” *2020 IEEE Students' Conference on Engineering and Systems, SCES 2020*, no. July, 2020, doi: 10.1109/SCES50439.2020.9236754.
- [12] I. N. Joharee, N. N. W. Nik Hashim, and N. S. Mohd Shah, “Sentiment Analysis and Text Classification for Depression Detection,” *Journal of Integrated and Advanced Engineering (JIAE)*, vol. 3, no. 1, pp. 65–78, 2023, doi: 10.51662/jiae.v3i1.86.
- [13] M. T. Uliniansyah *et al.*, “Twitter dataset on public sentiments towards biodiversity policy in Indonesia,” *Data Brief*, vol. 52, Feb. 2024, doi: 10.1016/j.dib.2023.109890.
- [14] N. AGUSTINA and C. N. IHSAN, “Pendekatan Ensemble Untuk Analisis Sentimen Covid19 Menggunakan Pengklasifikasi Soft Voting An Ensemble Approach For Covid19 Sentiment Analysis Using Soft Voting Classifier,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 10, no. 2, pp. 263–270, 2023, doi: 10.25126/jtiik.2023106215.
- [15] E. H. Muktafin and P. Kusriani, “Sentiments analysis of customer satisfaction in public services using K-nearest neighbors algorithm and natural language processing approach,” *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 1, pp. 146–154, Feb. 2021, doi: 10.12928/TELKOMNIKA.V19I1.17417.
- [16] M. A. Fauzi, “Random forest approach fo sentiment analysis in Indonesian language,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 1, pp. 46–50, 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.
- [17] X. Wan and J. Du, “Cloud classification for ground-based sky image using random forest,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 43, no. B3, pp. 835–842, 2020, doi: 10.5194/isprs-archives-XLIII-B3-2020-835-2020.
- [18] J. Khatib Sulaiman, D. Wijayanto, B. Pulu Hartato, and U. Amikom Yogyakarta, “Analisis Perbandingan Performa Algoritma XGBoost dan LightGBM pada Klasifikasi Kanker Payudara,” *Indonesian Journal of Computer Science*.