

## **ANALISIS SENTIMEN OBJEK WISATA DI PROVINSI SULAWESI SELATAN BERDASARKAN ULASAN PENGUNJUNG MENGGUNAKAN METODE *RANDOM FOREST CLASSIFIER***

Aisyah Nurul Izza<sup>\*1</sup>, Dian Eka Ratnawati<sup>2</sup>, Widhy Hayuhardhika Nugraha Putra<sup>3</sup>

<sup>1,2,3</sup>Universitas Brawijaya

Email: <sup>1</sup>aisyahh432@student.ub.ac.id, <sup>2</sup>dian\_ilkom@ub.ac.id, <sup>3</sup>widhy@ub.ac.id

<sup>\*</sup>Penulis Korespondensi

(Naskah masuk: 16 Februari 2022, diterima untuk diterbitkan: 27 April 2022)

### **Abstrak**

Provinsi Sulawesi Selatan merupakan salah satu gerbang bagi para pengunjung yang ingin berwisata di kawasan Indonesia Timur. Keindahan pantai dan objek wisata baharinya dapat mendukung perkembangan sektor pariwisata yang ada di Sulawesi Selatan. Objek wisata yang ada memiliki keunikan sehingga menjadi daya tarik bagi wisatawan lokal maupun mancanegara, salah satunya yaitu wisata pantai Tanjung Bira. Bagi wisatawan lokal dan mancanegara mencari referensi objek wisata melalui media sosial atau *platform* penyedia perjalanan. Pada penelitian ini penulis menggunakan website TripAdvisor yang merupakan website perjalanan wisata yang memungkinkan pengunjung memberikan ulasan terhadap suatu tempat, ulasan dari pengunjung tersebut akan dijadikan sebagai sumber data dalam penelitian ini. Tujuan dari penelitian ini adalah untuk mendapatkan informasi dari analisis sentimen melalui ulasan pengunjung objek wisata di Provinsi Sulawesi Selatan. Metode penelitian ini menggunakan algoritme *random forest classifier* dalam melakukan proses klasifikasi sentimen dengan mengklasifikasikan kedalam 3 kelas yaitu positif, netral, dan negatif. Proses klasifikasinya meliputi *text preprocessing*, pembobotan kata menggunakan metode TF-IDF dan klasifikasi menggunakan algoritme *random forest*. Hasil penelitian melalui pengklasifikasian ulasan pengunjung terhadap objek wisata di Provinsi Sulawesi Selatan sudah cukup baik untuk digunakan di dalam sistem dengan nilai *accuracy* sebesar 82%, *precision* sebesar 86% dan *recall* sebesar 86%.

**Kata kunci:** *random forest, decision tree, tf-idf, klasifikasi, objek wisata*

## ***SENTIMENT ANALYSIS OF TOURIST ATTRACTIONS IN THE PROVINCE OF SOUTH SULAWESI BASED ON VISITOR REVIEWS USING THE RANDOM FOREST CLASSIFICATION METHOD***

### ***Abstract***

*South Sulawesi Province is one of the gateways for tourists who want to travel in Eastern Indonesia. The beauty of the beach and the marine site could support the development of the tourism sector in South Sulawesi. The tourism site are unique so that they become an attraction for local and foreign tourists, one of that is Tanjung Bira beach. For local and international tourists who's looking for vacation spot references through social media or travel provider platforms. In this study, the author uses a platform called TripAdvisor as a source of research data. The purpose of this study was to obtain information from sentiment analysis*

*through visitor reviews of tourist attractions in South Sulawesi Province. This research method uses the Random Forest Classifier algorithm in conducting the sentiment classification process by classifying it into 3 classes, which are positive, neutral, and negative. The classification process includes text preprocessing, word weighting using the TF-IDF method and classification using the random forest algorithm. The results of the study through the classification of visitor reviews of tourist objects in South Sulawesi Province are good enough to be used in a system with accuracy of 82%, precision of 86% and recall of 86%.*

**Keywords:** *random forest, decision tree, TF-IDF, classification, tourist attraction*

---

## 1. PENDAHULUAN

Provinsi Sulawesi Selatan merupakan salah satu provinsi yang terletak di timur Indonesia yang terkenal akan keindahan wisata baharinya diantaranya Pantai Tanjung Bira, Wisata Alam Malino, Dan Taman Nasional Bantimurung. Sehingga Provinsi Sulawesi Selatan dapat dijadikan sebagai pilihan untuk menghabiskan liburan bersama keluarga. Melalui tagline “keSulselmi (ayo ke Sulsel)” pemerintah setempat Mencoba mengundang para wisatawan untuk berkunjung ke Sulawesi Selatan. Tetapi permasalahan yang dihadapi oleh Pemerintah Sulawesi Selatan dalam sektor pariwisatanya diantara yaitu karena terus meningkatnya jumlah wisatawan dan objek wisata dapat menyulitkan pemerintah setempat untuk memantau perkembangan objek wisata terbukti dari salah satu berita dari medcom.id bahwa “tempat wisata di Sulawesi Selatan belum didukung akses memadai”. Sehingga dibutuhkan ulasan pengunjung objek wisata untuk pengembangan dan pemeliharaan objek wisata yang ada. Dengan ulasan tersebut kita dapat menemukan sentimen apa yang diberikan pengunjung terhadap objek wisata yang ada.

Berdasarkan *Grafik Destination Analysis (2020)*, wisatawan pada zaman sekarang lebih memilih sumber dari internet untuk menentukan tujuan wisatanya. Ulasan dari internet merupakan hal yang penting karena dapat dijadikan sebagai informasi mengenai suatu tempat. Berbagai cara yang dapat dilakukan untuk penarikan ulasan dari pengunjung di suatu platform salah satunya adalah dengan menggunakan Analisis Sentimen. Analisis Sentimen merupakan pengolahan data untuk melacak opini masyarakat terhadap suatu jasa, produk atau topik tertentu dari internet. Dengan perkembangan teknologi informasi yang berbanding lurus dengan sektor pariwisata, sudah banyak platform yang dapat digunakan untuk memberikan ulasan suatu objek wisata yang akan peneliti gunakan yaitu website TripAdvisor.

Untuk melakukan analisis sentimen dibutuhkan metode, salah satunya adalah *random forest*. *Random forest* adalah algoritme pengklasifikasian yang terdiri dari beberapa *decision tree* yang dimana setiap cabang dibangun menggunakan vektor acak (Sandag, 2020). *Random forest* bersifat *ensemble* yaitu meningkatkan akurasi klasifikasi dengan menggabungkan beberapa metode klasifikasi. Kelebihan dari metode *random forest* adalah hasil akurasi yang bagus, mengatasi *noise* dan *missing value* serta cocok untuk mengklasifikasikan data dalam jumlah yang besar.

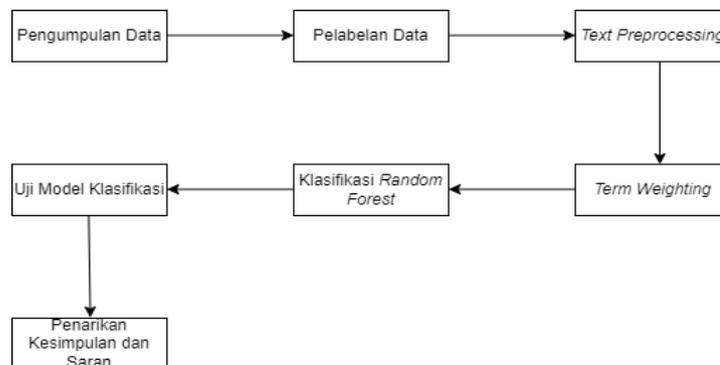
Penelitian ini akan dimulai dengan memberikan label sentimen positif, netral dan negatif secara manual dengan bantuan para ahli, *text preprocessing* yang meliputi *case folding* dan *tokenizing*, *filtering*, *stemming*, kemudian dilanjutkan dengan tahapan pembobotan kata menggunakan TF-IDF dan yang terakhir adalah evaluasi algoritme menggunakan metode pengujian *Confusion Matrix* dengan mempertimbangkan nilai *precision*, *recall*, *f1-score* dan *accuracy* yang dihasilkan. Melalui penelitian ini diharapkan mempermudah Dinas Pariwisata dan Kebudayaan Sulawesi Selatan untuk memantau perkembangan objek wisata di Sulawesi Selatan dan dengan adanya tanggapan dan ulasan dari pengunjung maka pemerintah setempat

dapat mengetahui langkah apa yang harus dilakukan dalam pengembangan objek wisata yang ada ke depannya.

Terdapat beberapa kajian penelitian terdahulu yang di anggap relevan pada penelitian ini yaitu penelitian oleh Baskoro (2021) telah menganalisis sentimen terhadap pelanggan hotel di Purwokerto dengan menggunakan metode *random forest* memiliki tingkat akurasi cukup tinggi sebesar 87,23% . Penelitian dari Fitri (2020) menunjukkan performa pengujian dengan menggunakan model algoritme *random forest* memiliki akurasi yang sangat tinggi sebesar 97,16% dibanding dengan beberapa model klasifikasi seperti naïve bayes sebesar 94,16% dan *support vector machine* (SVM) sebesar 96,01%. Febrianti (2020) telah menganalisis sentimen terhadap data ulasan pengunjung objek wisata Lawang Sewu kota Semarang pada situs TripAdvisor memiliki akurasi algoritme *random forest* sebesar 87,8% untuk ulasan Bahasa Inggris dan 80,0% untuk Bahasa Indonesia. Penelitian selanjutnya berasal dari Fauzi (2018) dengan judul analisis sentimen dalam Bahasa Indonesia dengan pendekatan *random forest* menunjukkan hasil akurasi sebesar 83,02%. Penelitian Setiowati & Helen (2018) yaitu klasifikasi analisis sentimen mengenai hotel di Yogyakarta menggunakan berbagai metode salah satunya adalah *random forest* yang menghasilkan akurasi sebesar 84,6%. Berdasarkan penelitian-penelitian sebelumnya algoritme *random forest* tergolong menghasilkan tingkat akurasi yang baik.

## 2. METODE PENELITIAN

Penelitian ini menggunakan algoritme *random forest* dalam melakukan proses klasifikasi sentimen terhadap objek wisata di Provinsi Sulawesi Selatan. Terdapat 8 tahapan metodologi penelitian yaitu dimulai dari studi literatur, pengumpulan data, pelabelan data, *text preprocessing*, pembobotan kata atau *term weighting*, klasifikasi *random forest*, analisis dan pengujian hasil klasifikasi, serta penarikan kesimpulan dan saran. Berikut diagram alur dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

### 2.1. Studi Literatur

Studi literatur adalah kegiatan mengkaji penelitian yang telah dilakukan sebelumnya dan teori-teori yang sesuai dalam proses penyelesaian masalah penelitian. Proses pengkajian penelitian nya dimulai dari mengumpulkan referensi dari buku dan jurnal yang serupa dengan teori maupun penelitian yang akan dilakukan. Teori tersebut meliputi analisis sentimen terhadap objek wisata, *data mining*, *text mining*, *text preprocessing*, pembobotan kata dengan *TF-IDF*, klasifikasi dengan metode *random forest*, serta pengujian dengan metode *Confusion Matrix* dan validasi data menggunakan *K-fold cross validation*.

## 2.2. Pengumpulan Data

Data yang dikumpulkan berupa ulasan pengunjung dengan Bahasa Indonesia terhadap objek wisata di Sulawesi Selatan melalui platform TripAdvisor. Teknik yang digunakan dalam pengumpulan data yaitu teknik Web Scraping dengan tools *Webscraper.io*. Data yang berhasil dikumpulkan secara keseluruhan yaitu sebanyak 1055 dengan Atribut data yang diambil meliputi Nama pengunjung dan teks ulasan.

## 2.3. Pelabelan Data

Proses pelabelan data dilakukan secara manual oleh para ahli yang berjumlah 3 orang ahli yang memiliki latar belakang pendidikan Bahasa Indonesia. Pelabelan dengan jumlah ganjil bertujuan untuk menghindari pendapat subjektif pada proses pelabelan. Teknik proses pelabelan seperti pada Tabel 1 yaitu memberikan label positif = 1, netral = 0 atau negatif = -1 melalui diskusi oleh para ahli yang bersangkutan.

Tabel 1. Pelabelan Data Manual

Ulasan	Ahli 1	Ahli 2	Ahli 3	Hasil
Mudah untuk mengakses, tempat yang bagus untuk kunjung dengan teman dan keluarga.	1	0	1	1

## 2.4. Text Preprocessing

Preprocessing adalah tahapan dimana data akan dibersihkan menggunakan beberapa tahap seperti *case folding* yaitu tahapan untuk mengubah seluruh huruf menjadi *lowercase*, *tokenizing* yaitu tahapan untuk memisahkan sebuah kalimat menjadi setiap kata yang ada di dalam kalimat, *Stopword removal* yaitu tahapan untuk menghilangkan kata yang tidak penting yang berada di dalam dokumen, dan *stemming* yaitu tahapan menghilangkan kata imbuhan dan mengubah kata jadi kata dasar (Khairunnisa et al., 2021). Pada Tabel 2 dapat dilihat data hasil dari *text preprocessing*.

Tabel 2. Hasil proses *text preprocessing*

Sebelum proses <i>text preprocessing</i>	Sesudah <i>text preprocessing</i>
Mudah untuk mengakses, tempat yang bagus untuk kunjung dengan teman dan keluarga.	['mudah', 'akses', 'bagus', 'kunjung', 'teman', 'keluarga']

## 2.5. Pembobotan Kata (Term Weighting)

Metode *TF-IDF* merupakan proses untuk memberikan bobot kata (*term*) terhadap dokumen (Nurjannah & Fitri Astuti, 2013). Memberikan pembobotan dilakukan dengan menghitung frekuensi kemunculan kata di dalam sebuah dokumen tersebut. Semakin besar munculnya suatu kata dalam dokumen Akan menunjukkan semakin umum kata tersebut sehingga bobot hubungan antara kata dan sebuah dokumen akan tinggi apabila frekuensi kata yang mengandung kata yang sama. Menurut (Yunus, 2020) menjelaskan, pembobotan ini nantinya Akan digunakan untuk algoritme klasifikasi dokumen.

1. *Term Frequency–Inverse Document Frequency (TF-IDF)* adalah bobot term terhadap dokumen. TF-IDF dirumuskan sebagai berikut.
2. *Term Frequency (TF)* adalah jumlah kemunculan term dalam dokumen. Pada *Term Frequency (TF)*, terdapat beberapa jenis formula yang dapat digunakan.

Rumus untuk TF-IDF dapat dilihat pada persamaan 1 dan persamaan 2 berikut.

$$tf = 0,5 + 0,5 \times \frac{tf}{\max(tf)} \quad (1)$$

$$idf_t = \log \frac{D}{df_t} \quad (2)$$

## 2.6. Random Forest Classifier

*Random Forest* adalah salah satu metode klasifikasi yang berkerja dengan cara membangun banyak pohon klasifikasi (Sandag, 2020). Metode ini dapat membangun simpul anak untuk di setiap *node* dan dilakukan pemilihan secara acak untuk memilah setiap simpul sehingga menghasilkan nilai akurasi relatif tinggi Menurut (Fitri, 2020).

Algoritme ini terdiri dari beberapa node yaitu *root node*, *internal node*, dan *leaf node*. *Root node* biasa disebut juga sebagai akar dari pembentukan *tree*, *internal node* biasa disebut juga *node* percabangan dan *leaf node* adalah *node* terakhir yang terbentuk. Perhitungan *random forest* kurang lebih sama dengan perhitungan untuk *decision tree* dimulai dengan menghitung tingkat ketidakmurnian atribut dapat menggunakan *entropy* atau *gini impurity*. Untuk penelitian ini menggunakan *gini impurity* untuk menghitung ketidakmurnian atribut. Rumus untuk menghitung *gini impurity* dapat dilihat pada persamaan 3, sedangkan rumus untuk *information gain* dapat dilihat pada persamaan 4.

$$G = 1 - \sum_i^n P_i^2 \quad (3)$$

$$\text{Information Gain} = \text{Gini Impurity} - \text{Gini Average} \quad (4)$$

## 2.7. Confusion Matrix

*Confusion matrix* adalah metode pengujian yang digunakan untuk menghitung akurasi suatu algoritme data mining (Sandag, 2020). Evaluasi *confusion matrix* ialah memprediksi *true* atau *false* dari objek untuk memberikan nilai akurasi nya. *Precision* adalah tingkat ketepatan untuk prediksi benar positif terhadap keseluruhan data hasil prediksi positif. *Recall* adalah tingkat ketepatan ratio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. *Confusion matrix* menguji dengan memberikan akurasi terhadap *precision*, *recall*, *f1-score* dan *accuracy* dengan mempertimbangkan *True Positive*, *True Negative*, *False Positive*, dan *False Negative*.

## 3. HASIL DAN PEMBAHASAN

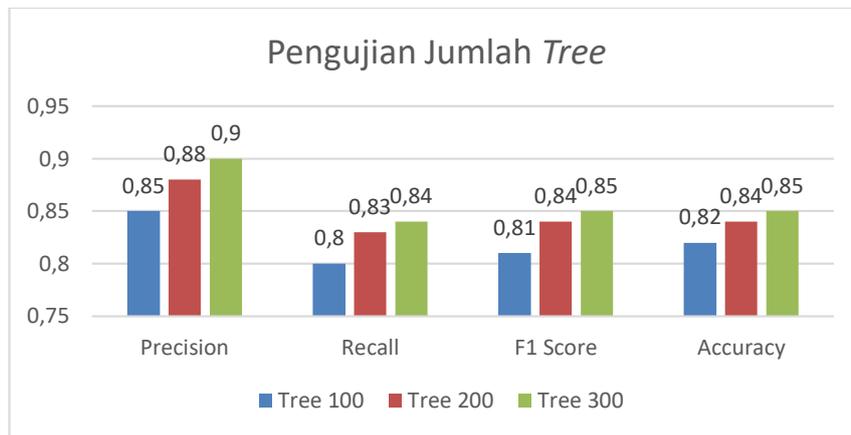
Penelitian ini terdapat 3 parameter pengujian yang akan diuji yaitu: pengujian jumlah *tree*, pengujian kedalaman *tree*, pengujian jumlah data, dan pengujian algoritme *random forest* menggunakan *k-fold cross validation*. Pengujian ini menggunakan metode *confusion matrix* yang akan memberikan keluaran nilai *accuracy*, *precision*, *recall* dan *f1-score*.

### 3.1. Pegujian Jumlah Tree

Pengujian ini bertujuan untuk mengetahui pengaruh jumlah *tree* dengan performa terbaik dan bagaimana pengaruh perubahan jumlah *tree* terhadap akurasi. Pengujian menggunakan proporsi data sebanyak 90% data latih dan 10% data uji, jumlah *tree* yang diujikan sebesar 100, 200, dan 300, menggunakan kedalaman *tree* sebesar 50. Data yang digunakan sebelumnya telah melewati proses text preprocessing. Tabel 3 dan Gambar 2 memperlihatkan hasil pengaruh jumlah *tree* terhadap akurasi akhir.

Tabel 3. Hasil Pengujian Jumlah *Tree*

Jumlah Tree	Jumlah Kedalaman Tree	Precision	Recall	F1-Score	Accuracy
100	50	0.80	0.81	0.81	0.82
200	50	0.88	0.83	0.84	0.84
300	50	0.90	0.84	0.85	0.85



Gambar 2. Pengujian Jumlah *Tree*

Pada Gambar 2 dapat dilihat bahwa rata-rata akurasi tertinggi pada pengujian terdapat pada jumlah *tree* sebesar 300 dengan rata-rata akurasi sebesar 85%. Dari analisis ini dapat disimpulkan bahwa jumlah *tree* berpengaruh terhadap akurasi karena semakin banyak *tree* terbentuk maka semakin tinggi nilai akurasi dan semakin terperinci *tree* dalam mengklasifikasikan suatu dataset.

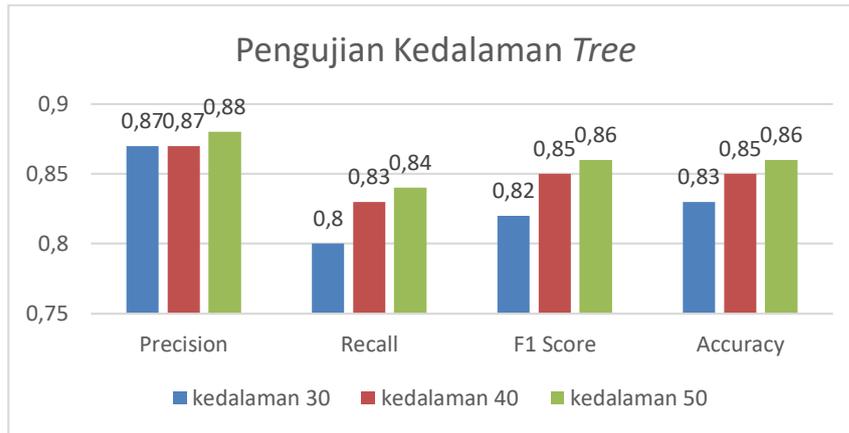
### 3.2. Pengujian Jumlah Kedalaman *Tree*

Pengujian ini bertujuan untuk mengetahui pengaruh jumlah kedalaman *tree* dengan performa terbaik dan bagaimana pengaruh perubahan jumlah kedalaman *tree* terhadap akurasi. Pengujian menggunakan jumlah kedalaman *tree* yang dipilih secara acak sebesar 30, 40, dan 50 dengan jumlah *tree* sebesar 300 kemudian menggunakan data latih sebanyak 90% dan data uji 10%. Data yang digunakan sebelumnya telah melewati proses text preprocessing. Tabel 4 dan Gambar 3 akan memperlihatkan hasil pengaruh kedalaman *tree* terhadap akurasi akhir.

Pada Gambar 3 dapat dilihat bahwa rata-rata akurasi tertinggi pada pengujian terdapat pada jumlah kedalaman *tree* sebesar 50 dengan rata-rata akurasi sebesar 86% hal ini disebabkan karena semakin banyak kedalaman *tree* dalam pengklasifikasian maka semakin banyak proses *splitting data*. *Splitting data* merupakan proses yang dilakukan untuk mendapatkan aturan baru berisikan informasi mengenai data yang digunakan.

Tabel 4. Hasil Pengujian Kedalaman *Tree*

Jumlah Tree	Kedalaman Tree	Precision	Recall	F1-Score	Accuracy
300	30	0.87	0.80	0.82	0.83
300	40	0.87	0.83	0.85	0.85
300	50	0.88	0.84	0.86	0.86



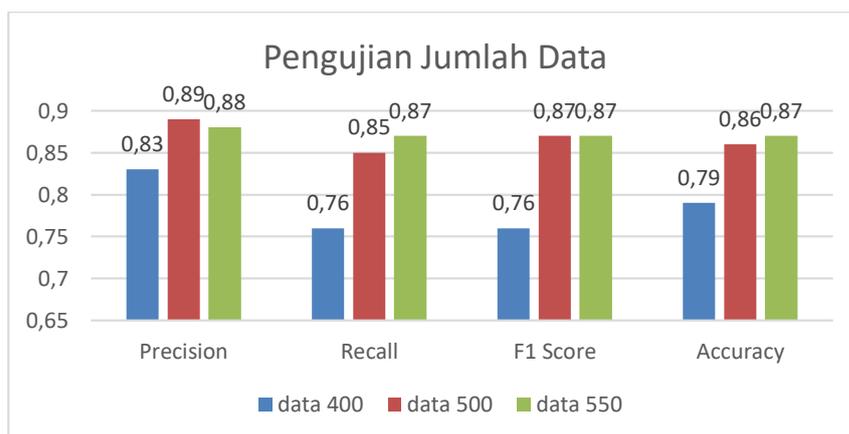
Gambar 3. Pengujian Kedalaman Tree

### 3.3. Pengujian Jumlah Data

Pengujian jumlah data bertujuan untuk mengetahui hubungan jumlah data positif, netral, dan negatif dengan hasil akurasi. Pada proses pengujian dilakukan 3 kali yaitu jumlah data sebanyak 400, 500, dan 550. Data yang digunakan sebelumnya telah melewati proses *text preprocessing*. Pengujian yang akan dijalankan sebanyak 5 kali kemudian hasil pengujian akan diambil rata-ratanya untuk hasil akhir akurasi. Tabel 5 dan Gambar 4 akan memperlihatkan hasil pengaruh jumlah data terhadap akurasi akhir.

Tabel 5. Hasil Pengujian Jumlah Data

Jumlah Data	Jumlah Tree	Kedalaman Tree	Precision	Recall	F1-Score	Accuracy
400	300	50	0.83	0.76	0.76	0.79
500	300	50	0.89	0.85	0.87	0.86
550	300	50	0.88	0.87	0.87	0.87



Gambar 4. Pengujian Jumlah Data

Pada Gambar 4 dapat dilihat bahwa rata-rata akurasi tertinggi pada pengujian terdapat pada jumlah data sebesar 550 dengan rata-rata akurasi sebesar 87%. Hal ini disebabkan jumlah data yang digunakan harus sesuai dengan kebutuhan.

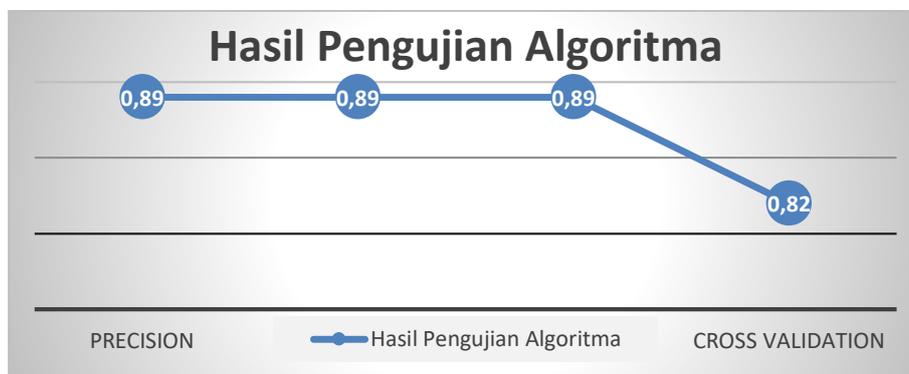
### 3.4. Pengujian Algoritme Random Forest

Pengujian ini bertujuan untuk mengetahui performa algoritme *random forest* untuk klasifikasi dan hasil akurasi algoritme. Pada proses pengujian menggunakan *cross validation* yang dilakukan dengan menggabungkan parameter terbaik yang telah diuji sebelumnya.

Pengujian yang akan dijalankan sebanyak 5 *fold* kemudian hasil pengujian akan diambil rata-ratanya untuk hasil akhir akurasi. Tabel 6 dan Gambar 5 akan memperlihatkan hasil akurasi algoritme *random forest* dengan beberapa parameter pengujian sebelumnya.

Tabel 6. Hasil Pengujian Algoritme *Random Forest*

K-Fold ke-	Precision	Recall	F1-Score	Accuracy
1	0.92	0.91	0.91	0.83
2	0.85	0.86	0.85	0.83
3	0.92	0.92	0.92	0.81
4	0.86	0.86	0.86	0.82
5	0.91	0.91	0.91	0.83
<b>Total Pengujian</b>	<b>0.89</b>	<b>0.89</b>	<b>0.89</b>	<b>0.82</b>



Gambar 5. Pengujian Algoritme *Random Forest*

Berdasarkan Gambar 5 dapat dilihat hasil pengujian dari *K-fold cross validation* bahwa algoritme *random forest* memiliki tingkat akurasi yang cukup baik. Hal ini disebabkan data latih yang diambil secara *random* dari data original. Selain itu, *random forest* dapat membentuk jumlah *tree* sesuai dengan kebutuhan.

#### 4. KESIMPULAN

Metode *random forest* dapat diimplementasikan pada proses analisis sentimen objek wisata di Provinsi Sulawesi Selatan menggunakan data dari ulasan-ulasan yang diberikan oleh pengunjung pada website TripAdvisor. Data yang digunakan harus dilabeli dengan kelas sentimen positif, netral atau negatif. Proses awal klasifikasi sentimen pada penelitian ini yaitu memberikan pelabelan data terhadap ulasan pengunjung, pelabelan ini meliputi positif, netral dan negatif. Setelah pelabelan, selanjutnya adalah proses *text preprocessing* yang meliputi tahapan *case folding*, *tokenizing*, *Stopword Removal*, dan *stemming*. Setelah proses *text preprocessing* selanjutnya adalah pembobotan kata menggunakan metode *TF-IDF* dan dilanjutkan untuk pengujian algoritme *random forest* menggunakan *k-fold cross validation*.

Dari hasil pengujian yang dilakukan di dalam penelitian dapat menunjukkan beberapa hal sebagai berikut:

- Dari analisis ini dapat disimpulkan bahwa jumlah *tree* berpengaruh terhadap akurasi karena semakin banyak *tree* yang terbentuk maka semakin tinggi nilai akurasi dan semakin terperinci *tree* dalam mengklasifikasikan suatu dataset.
- Semakin banyak kedalaman *tree* dalam pengklasifikasian maka semakin banyak proses *splitting data*. *Splitting data* merupakan proses yang dilakukan untuk mendapatkan aturan baru berisikan informasi mengenai data yang digunakan. Jadi semakin banyak informasi yang digunakan maka semakin baik hasil yang didapatkan.

- c. Semakin banyak data yang digunakan maka akurasi yang didapatkan juga semakin tinggi tetapi tergantung dengan jumlah data yang digunakan harus sesuai dengan kebutuhan
- d. Penerapan metode klasifikasi *random forest* dalam mengklasifikasikan data ulasan pengunjung mengenai objek wisata yang ada di Provinsi Sulawesi Selatan menghasilkan *accuracy* sebesar 82%, nilai *precision* sebesar 86%, dan nilai *recall* sebesar 86%. Dari nilai akurasi yang didapatkan menunjukkan bahwa performa algoritme *random forest* dalam klasifikasi ulasan pengunjung terhadap objek wisata di Provinsi Sulawesi Selatan sudah cukup baik untuk digunakan di dalam sistem.

Dalam menyempurnakan penelitian ini selanjutnya disarankan untuk Membandingkan dengan algoritme klasifikasi yang lain untuk menemukan algoritme klasifikasi yang terbaik dan menambahkan data ulasan menggunakan bahasa inggris atau bahasa asing yang lainnya.

## DAFTAR PUSTAKA

- BASKORO., dkk. 2021. *Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR)*. 8106, 21–29. <https://doi.org/10.20895/INISTA.V3>
- FAUZI, M. A. 2018. Random forest approach fo sentiment analysis in Indonesian language. *Indonesian Journal of Electrical Engineering and Computer Science*, 12(1), 46–50. <https://doi.org/10.11591/ijeecs.v12.i1.pp46-50>
- FEBRIANTI, A. V. 2020. *Analisis Sentimen Data Ulasan Pengunjung Objek Wisata Lawang Sewu Kota Semarang Pada Situs Tripadvisor*. <http://lib.unnes.ac.id/41832/1/4112317002.pdf>
- FITRI, E. 2020. Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. *Jurnal Transformatika*, 18(1), 71. <https://doi.org/10.26623/transformatika.v18i1.2317>
- KHAIRUNNISA, S., ADIWIJAYA, A., & FARABY, S. AL. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *Jurnal Media Informatika Budidarma*, 5(2), 406. <https://doi.org/10.30865/mib.v5i2.2835>
- NURJANNAH, M., & FITRI ASTUTI, I. 2013. Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) Untuk Text Mining Mahasiswa S1 Program Studi Ilmu Komputer FMIPA Universitas Mulawarman Dosen Program Studi Ilmu Komputer FMIPA Universitas Mulawarman. *Jurnal Informatika Mulawarman*, 8(3), 110–113.
- SANDAG, G. A. 2020. Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest. *CogITo Smart Journal*, 6(2), 167. <https://doi.org/10.31154/cogito.v6i2.270.167-178>
- SETIOWATI, Y., & HELEN, A. 2018. Klasifikasi Analisis Sentimen Mengenai Hotel Di Yogyakarta. *SCAN - Jurnal Teknologi Informasi Dan Komunikasi*, 13(1). <https://doi.org/10.33005/scan.v13i1.1052>
- HERDHianto, A., dkk. 2020. *Sentiment Analysis Menggunakan Naïve Bayes Classifier (Nbc ) Pada Tweet Tentang Zakat Sentiment Analysis Menggunakan Naïve Bayes Classifier ( Nbc ) Pada Tweet Tentang*.
- ROZI, I., PRAMONO, S., & DAHLAN, E. 2012. Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi. *Jurnal EECCIS*, 6(1), 37–43.