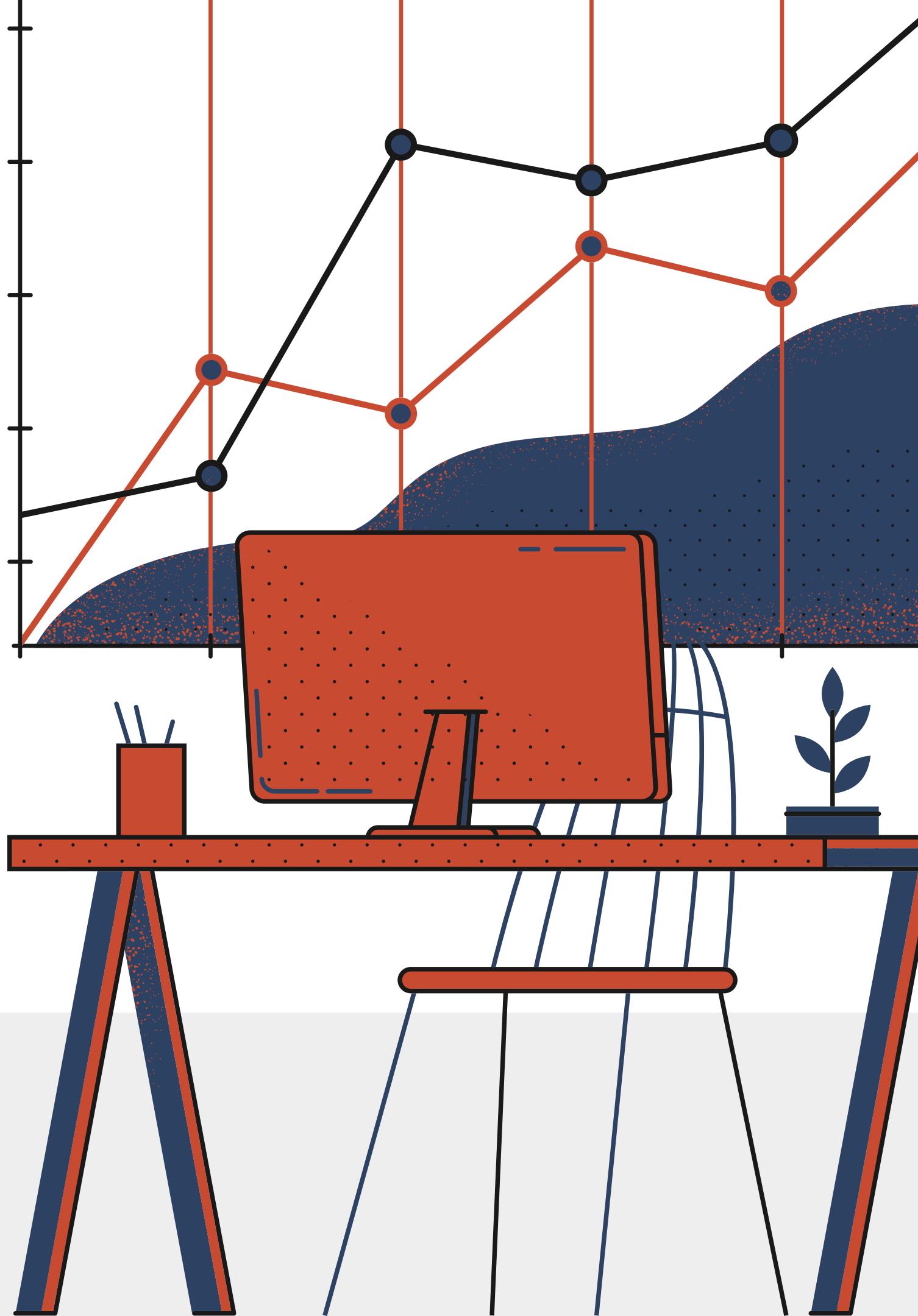


# **Analisis Sentimen Persepsi Masyarakat Tentang Wisata Pantai Sembilan Menggunakan Random Forest**

**Mitha Putri Oktaviana**



[linkedin.com/mithaaapr](https://linkedin.com/mithaaapr)



# Latar Belakang

1. Pantai Sembilan yang cukup terkenal di Kabupaten Sumenep. Namun walaupun cukup terkenal dengan keindahan pantainya, masih terdapat beberapa kekurangan dari berbagai aspek seperti kebersihan, keamanan dan keindahan. Hal ini dapat dilihat pada ulasan yang diberikan masyarakat melalui Google Maps.
2. Digunakan analisis sentimen dengan tujuan dapat menjadi bahan acuan untuk pengembangan wisata ini.
3. Pada penelitian, data akan diolah dan dikategorikan menjadi 3 kelas yaitu positif, negatif, dan netral. Yang kemudian akan diklasifikasikan ke dalam algoritma Random Forest.





## Pertanyaan Penelitian

1. Bagaimana pengaruh jumlah dan kedalaman tree terhadap hasil penelitian?
2. Bagaimana performa accuracy, recall, precision, dan F1-Score yang didapatkan menggunakan metode Random Forest Classifier?

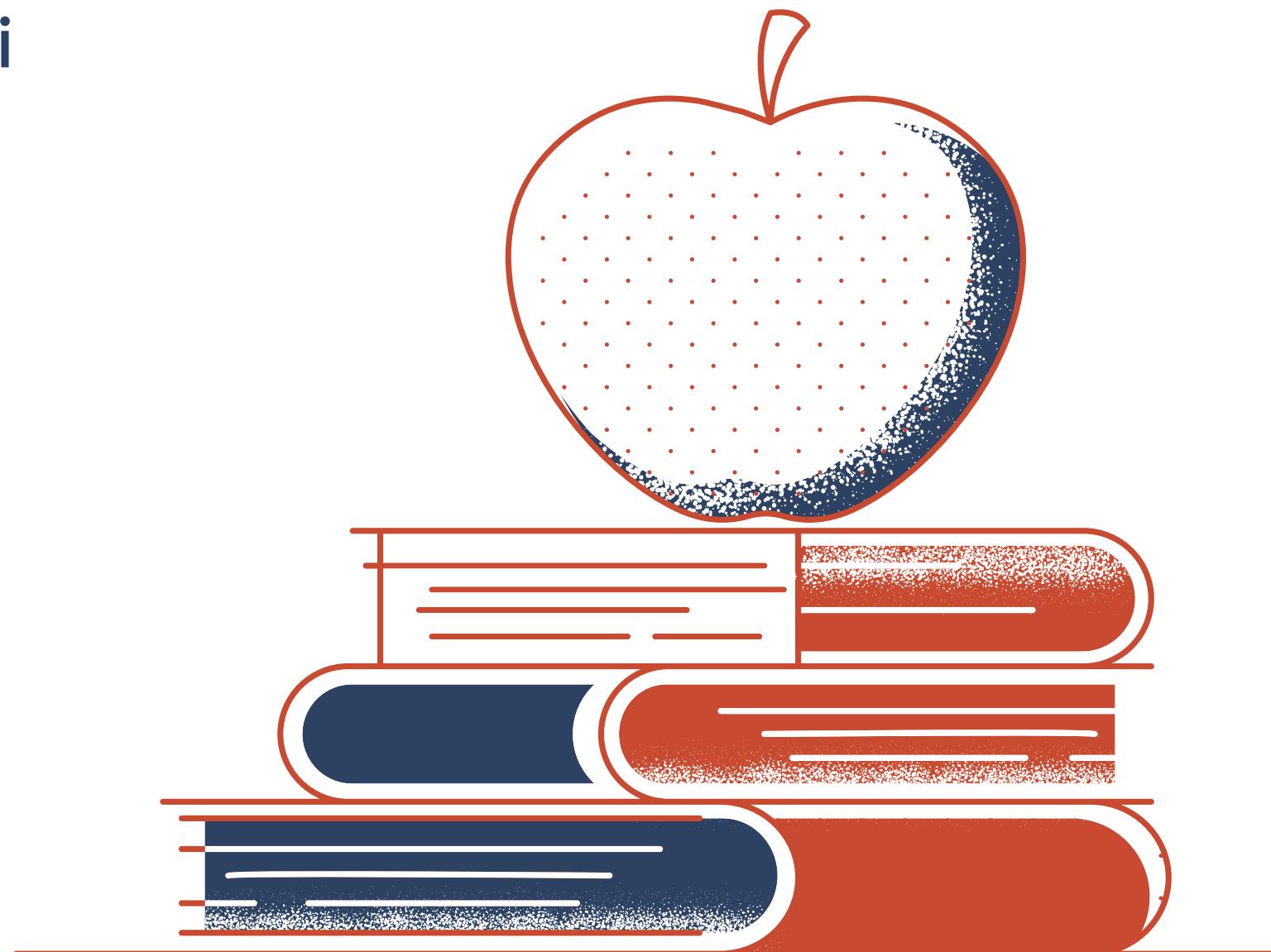
# Batasan Masalah

1. Data yang diambil adalah dari Ulasan Masyarakat pada Google Maps.
2. Data yang diambil menggunakan Teknik Scrapping dengan memasukkan url dari website dengan kata kunci “Pantai Sembilan Gmaps”.
3. Data yang berhasil diambil adalah sebanyak 1.981 data.
4. Data yang diambil adalah ulasan dalam bentuk Bahasa Indonesia.
5. Klasifikasi terdiri dari 3 kelas yaitu positif, negatif, dan netral.
6. Perhitungan TF-IDF, dan Random Forest Classifier dilakukan dengan bantuan library python.

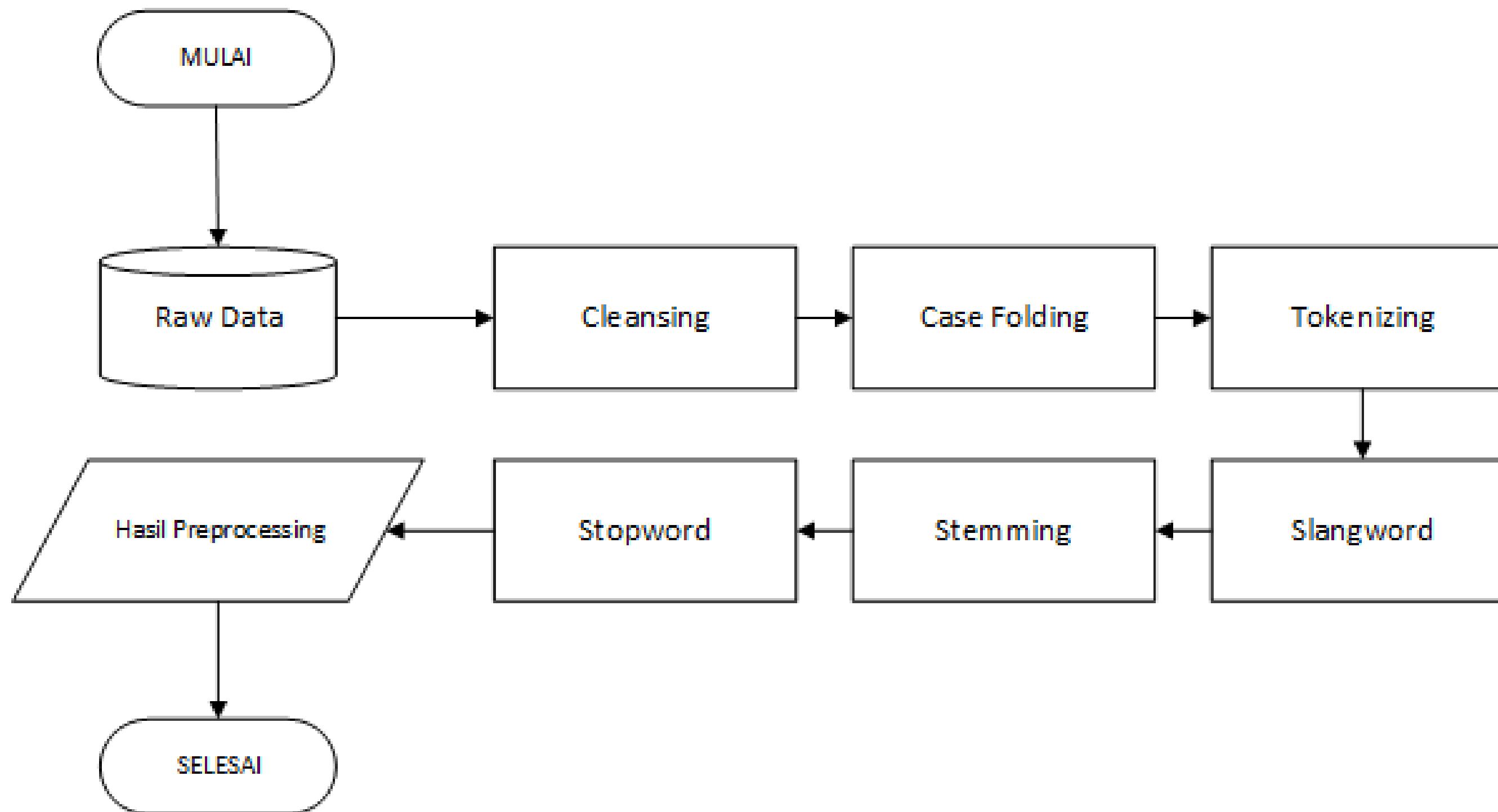


# Pengumpulan Data

- Data didapatkan dari ulasan masyarakat dari Google Maps.
- Proses Scrapping Data menggunakan Tools WebHarvy, dengan atribut yang diambil diantaranya nama, waktu, rating, dan review.
- Pada proses scrapping didapatkan sebanyak 1.981 data dengan rentang tahun 2017-2022.



# Preprocessing



## Cleansing

Teks Review	Sesudah tahap Cleansing
Pantai dengan ombak yang tenang dan cukup murah tiket masuk maupun penyeberangannya. Pasir putih dan tersedia penginapan tepi pantai yang murah dan cukup bagus. Perlu ditingkatkan kebersihan tepi pantainya oleh pengelola. Semoga tetap asri dan berkembang maju	Pantai dengan ombak yang tenang dan cukup murah tiket masuk maupun penyeberangannya Pasir putih dan tersedia penginapan tepi pantai yang murah dan cukup bagus Perlu ditingkatkan kebersihan tepi pantainya oleh pengelola Semoga tetap asri dan berkembang maju

## Case Folding

Teks Review Sesudah tahap Cleansing	Sesudah tahap Case Folding
Pantai dengan ombak yang tenang dan cukup murah tiket masuk maupun penyeberangannya Pasir putih dan tersedia penginapan tepi pantai yang murah dan cukup bagus Perlu ditingkatkan kebersihan tepi pantainya oleh pengelola Semoga tetap asri dan berkembang maju	pantai dengan ombak yang tenang dan cukup murah tiket masuk maupun penyeberangannya pasir putih dan tersedia penginapan tepi pantai yang murah dan cukup bagus perlu ditingkatkan kebersihan tepi pantainya oleh pengelola semoga tetap asri dan berkembang maju

## Tokenizing

Teks Review Sesudah tahap Case Folding	Sesudah tahap Tokenizing
pantai dengan ombak yang tenang dan cukup murah tiket masuk maupun penyeberangannya pasir putih dan tersedia penginapan tepi pantai yang murah dan cukup bagus perlu ditingkatkan kebersihan tepi pantainya oleh pengelola semoga tetap asri dan berkembang maju	['pantai', 'dengan', 'ombak', 'yang', 'tenang', 'dan', 'cukup', 'murah', 'tiket', 'masuk', 'maupun', 'penyeberangannya', 'pasir', 'putih', 'dan', 'tersedia', 'penginapan', 'tepi', 'pantai', 'yang', 'murah', 'dan', 'cukup', 'bagus', 'perlu', 'dingkatkan', 'kebersihan', 'tepi', 'pantainya', 'oleh', 'pengelola', 'semoga', 'tetap', 'asri', 'dan', 'berkembang', 'maju']

## Slangword

Teks Review Sesudah tahap Tokenizing	Sesudah tahap Konversi Slangword
['pantai', 'dengan', 'ombak', 'yang', 'tenang', 'dan', 'cukup', 'murah', 'tiket', 'masuk', 'maupun', 'penyeberangannya', 'pasir', 'putih', 'dan', 'tersedia', 'penginapan', 'tepi', 'pantai', 'yang', 'murah', 'dan', 'cukup', 'bagus', 'perlu', 'dingkatkan', 'kebersihan', 'tepi', 'pantainya', 'oleh', 'pengelola', 'semoga', 'tetap', 'asri', 'dan', 'berkembang', 'maju']	['pantai', 'dengan', 'ombak', 'yang', 'tenang', 'dan', 'cukup', 'murah', 'tiket', 'masuk', 'maupun', 'penyeberangannya', 'pasir', 'putih', 'dan', 'tersedia', 'penginapan', 'tepi', 'pantai', 'yang', 'murah', 'dan', 'cukup', 'bagus', 'perlu', 'dingkatkan', 'kebersihan', 'tepi', 'pantainya', 'oleh', 'pengelola', 'semoga', 'tetap', 'asri', 'dan', 'berkembang', 'maju']

## Stemming

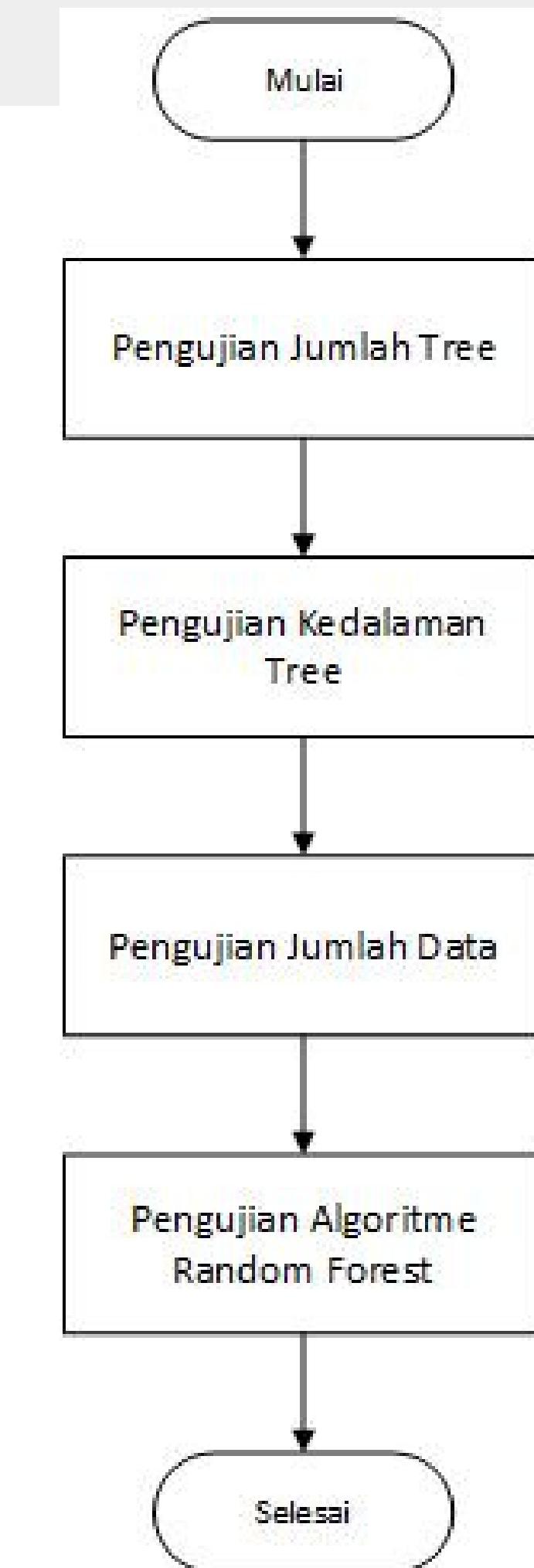
Teks Review Sesudah tahap Konversi Slangword	Sesudah tahap Stemming
['pantai', 'dengan', 'ombak', 'yang', 'tenang', 'dan', 'cukup', 'murah', 'tiket', 'masuk', 'maupun', 'penyeberangannya', 'pasir', 'putih', 'dan', 'tersedia', 'penginapan', 'tepi', 'pantai', 'yang', 'murah', 'dan', 'cukup', 'bagus', 'perlu', 'dingkatkan', 'kebersihan', 'tepi', 'pantainya', 'oleh', 'pengelola', 'semoga', 'tetap', 'asri', 'dan', 'berkembang', 'maju']	['pantai', 'dengan', 'ombak', 'yang', 'tenang', 'dan', 'cukup', 'murah', 'tiket', 'masuk', 'maupun', 'seberang', 'pasir', 'putih', 'dan', 'sedia', 'inap', 'tepi', 'pantai', 'yang', 'murah', 'dan', 'cukup', 'bagus', 'perlu', 'tingkat', 'bersih', 'tepi', 'pantai', 'oleh', 'kelola', 'moga', 'tetap', 'asri', 'dan', 'kembang', 'maju']

## Stopword

Teks Review Sesudah tahap Stemming	Sesudah tahap Stopword
['pantai', 'dengan', 'ombak', 'yang', 'tenang', 'dan', 'cukup', 'murah', 'tiket', 'masuk', 'maupun', 'seberang', 'pasir', 'putih', 'dan', 'sedia', 'inap', 'tepi', 'pantai', 'yang', 'murah', 'dan', 'cukup', 'bagus', 'perlu', 'tingkat', 'bersih', 'tepi', 'pantai', 'oleh', 'kelola', 'moga', 'tetap', 'asri', 'dan', 'kembang', 'maju']	['pantai', 'ombak', 'tenang', 'murah', 'tiket', 'masuk', 'seberang', 'pasir', 'putih', 'sedia', 'inap', 'tepi', 'pantai', 'murah', 'bagus', 'tingkat', 'bersih', 'tepi', 'pantai', 'kelola', 'moga', 'asri', 'kembang', 'maju']

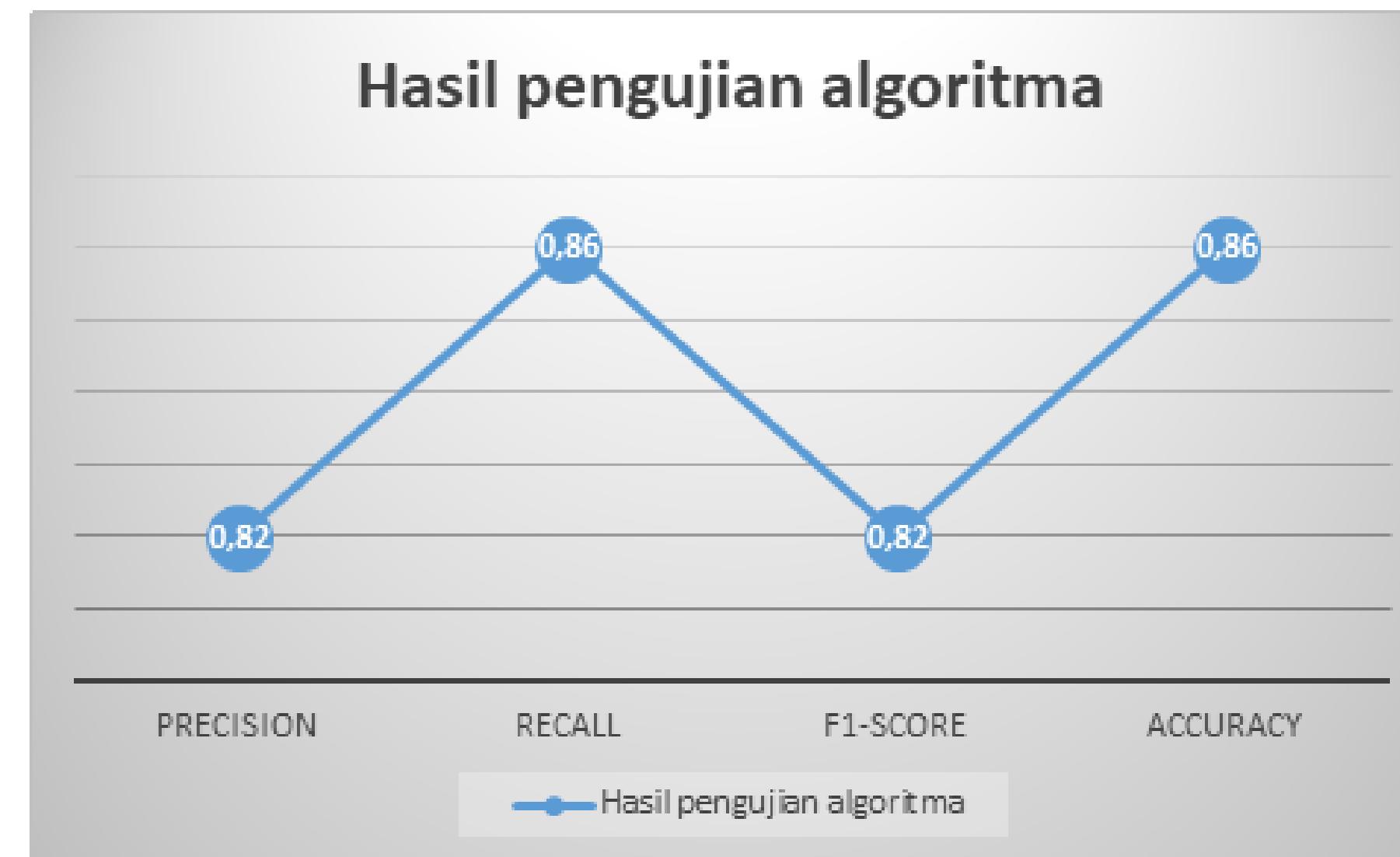
# Pengujian Algoritme Random Forest

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.



# Hasil dan Pembahasan

Dari analisis yang dilakukan pada penelitian ini menghasilkan nilai akurasi sebesar 86%



# Skenario Pengujian



- 1 Pengujian Jumlah Tree. Jumlah tree yang akan diujikan adalah sebesar 100, 200, dan 300, dan menggunakan kedalaman tree sebesar 50
- 2 Pengujian jumlah kedalaman tree. Pada penelitian ini, jumlah kedalaman tree yang akan diujikan diantaranya sebesar 30, 40, 50, dengan jumlah tree yang akan diujikan sebesar 300
- 3 Pengujian jumlah data. Proses pengujian pada penelitian ini dilakukan sebanyak 3 kali yaitu dengan jumlah data sebanyak, 400, 500, dan 550
- 4 Pengujian Klasifikasi Random Forest Classifier

# Hasil pengujian 1

Berikut tabel dari hasil pengujian 1 (Jumlah Tree) :

Jumlah Tree	Jumlah Kedalaman Tree	Precision	Recall	F1-Score	Accuracy
100	50	0,91	0,90	0,88	0,90
200	50	0,90	0,89	0,87	0,90
300	50	0,91	0,90	0,88	0,92

## Hasil pengujian 2

Berikut tabel dari hasil pengujian 2 (Jumlah kedalaman tree) :

Jumlah Tree	Jumlah Kedalaman Tree	Precision	Recall	F1-Score	Accuracy
300	30	0,90	0,85	0,79	0,85
300	40	0,87	0,87	0,83	0,87
300	50	0,90	0,90	0,8	0,90

# Hasil pengujian 3

Berikut tabel dari hasil pengujian 3 (Jumlah data) :

Jumlah Data	Jumlah Tree	Jumlah Kedalaman Tree	Precision	Recall	F1-Score	Accuracy
400	300	50	0,79	0,87	0,82	0,86
500	300	50	0,80	0,86	0,81	0,86
550	300	50	0,82	0,86	0,82	0,87

# Hasil pengujian 4

Berikut tabel dari hasil pengujian 4 (Random forest) :

K-Fold ke-	Precision	Recall	F1-Score	Accuracy
1	0,91	0,90	0,88	0,90
2	0,90	0,89	0,87	0,90
3	0,91	0,90	0,88	0,92
4	0,93	0,93	0,91	0,93
5	0,86	0,84	0,79	0,84
6	0,85	0,89	0,86	0,89
7	0,7	0,78	0,7	0,78
8	0,86	0,87	0,84	0,87
9	0,9	0,89	0,86	0,89
10	0,83	0,91	0,87	0,91
<b>Total Pengujian</b>	<b>0,82</b>	<b>0,86</b>	<b>0,82</b>	<b>0,86</b>

# Kesimpulan

1. Dari hasil analisis dapat disimpulkan bahwa jumlah tree sangat berpengaruh terhadap akurasi dikarenakan semakin banyak tree yang terbentuk maka semakin tinggi nilai akurasi dan semakin terperinci tree dalam mengklasifikasikan suatu dataset. Dan semakin banyak kedalaman tree maka semakin baik hasil yang didapatkan
2. Pada penelitian Analisis Sentimen Persepsi Masyarakat Tentang Pantai Sembilan Menggunakan Random Forest ini menghasilkan nilai **Accuracy sebesar 86%, precision 82%, recall 86% dan F1-Score 82%**. Dari nilai akurasi yang didapatkan menunjukkan bahwa performa Algoritme Random Forest ini sudah cukup baik untuk digunakan di dalam sistem.



# Terima Kasih



[linkedin.com/mithaaptr](https://linkedin.com/mithaaptr)