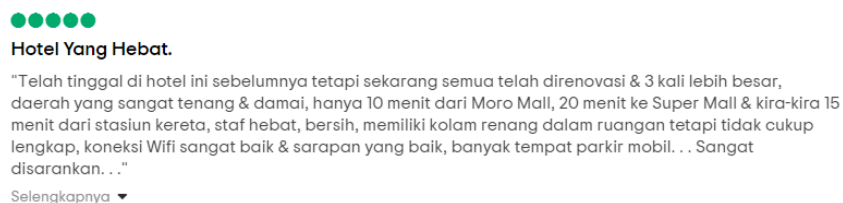


BAB III

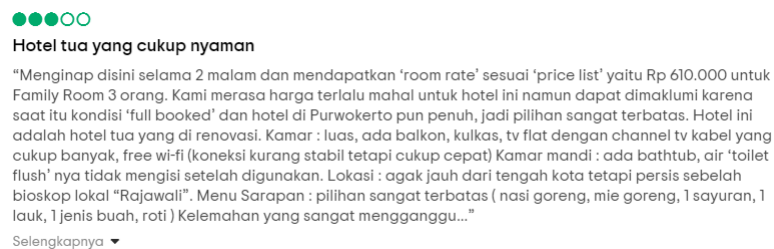
METODOLOGI PENELITIAN

3.1 Subjek dan Objek Penelitian

Subjek pada penelitian ini adalah para pelanggan hotel di Purwokerto. Pelanggan hotel biasanya menulis ulasan dikolom komentar setelah selesai menginap untuk membagikan pengalaman mereka. Ulasan tersebut dikelompokkan dalam menu pelanggan sebagai wisatawan bisnis, pasangan, wisatawan solo, keluarga dengan anak kecil, keluarga dengan anak dewasa dan rombongan. Berikut adalah contoh ulasan pelanggan hotel di Purwokerto pada situs [tripadvisor.co.id](https://www.tripadvisor.co.id) :

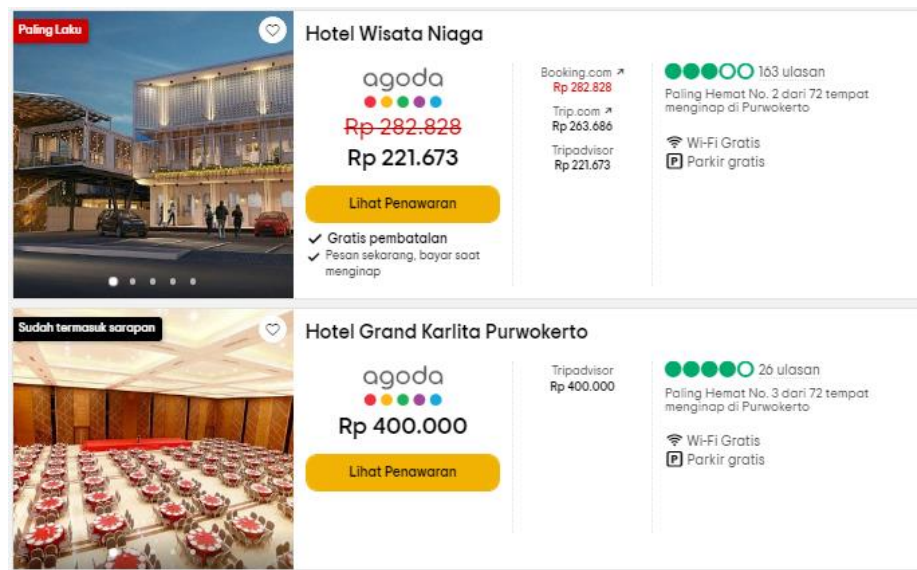


Gambar 3.1 Contoh ulasan pelanggan hotel TripAdvisor 1



Gambar 3.2 Contoh ulasan pelanggan hotel TripAdvisor 2

Objek pada penelitian ini adalah hotel di Purwokerto. Hotel di Purwokerto yang terdaftar dalam situs TripAdvisor. Hotel tersebut lebih tepatnya yang berada di wilayah Purwokerto. Dalam penelitian ini menggunakan data pelanggan dari hotel di Purwokerto yang ada di [tripadvisor.com](https://www.tripadvisor.com). Berikut adalah contoh tampilan hotel yang ada di Purwokerto yang terdaftar dalam situs [tripadvisor.com](https://www.tripadvisor.com) :



Gambar 3.3 Contoh tampilan hotel Purwokerto di TripAdvisor.co.id

3.2 Alat dan Bahan Penelitian

Alat yang digunakan dalam penelitian ini berupa *software* dan *hardware* yang dilengkapi dengan perangkat keras dan perangkat lunak pendukung. Sedangkan bahan yang digunakan adalah data ulasan pelanggan hotel melalui situs TripAdvisor.co.id.

3.2.1 Alat Penelitian

Dalam penelitian ini penulis menggunakan perangkat keras dengan spesifikasi sebagai berikut :

1. Laptop : ACER Predator Nitro 5.
2. Processor : Intel i7-7700HQ.
3. Kartu Grafis : Kartu grafis NVIDIA GeForce® GTX 1050.
4. Layar : 15,6 “full HD(1080)
5. Memory : RAM 8GB.
6. Hard Disk : 1TB HDD.

Adapun perangkat lunak sebagai kebutuhan *software* yang digunakan dalam penelitian ini sebagai berikut :

1. Sistem operasi : Windows 10 pro

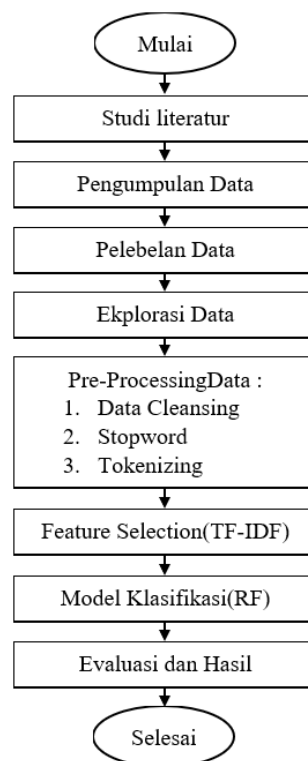
2. Webharvy
3. Microsoft Excel 2019
4. Python versi 3.7
 - a) Anaconda 3
 - b) Jupyter Notebook

3.2.2 Bahan Penelitian

Adapun bahan utama dari penelitian ini menggunakan data ulasan pelanggan yang diperoleh dari Situs TripAdvisor.co.id.

3.3 Diagram Alir Penelitian/Proses Penelitian

Pada penyusunan laporan penelitian ini terdapat beberapa tahapan dalam melakukan penelitian dan dalam pembuatan model analisis sentimen. Berikut merupakan tahapan yang dilakukan pada penyusunan laporan ini pada Gambar 3.4 dibawah ini :



Gambar 3.4 Diagram alir penelitian

3.3.1 Studi Literatur

Dalam penelitian ini dibutuhkan referensi sebagai dasar dalam melakukan tahapan penelitian, dan referensi yang digunakan oleh penulis menggunakan jurnal terdahulu yang masih berhubungan dengan penelitian ini sebagai referensi. Selain menggunakan jurnal terdahulu, penulis juga menggunakan buku untuk memperluas pengetahuan dalam memahami algoritma yang digunakan.

3.3.2 Pengumpulan Data

Pengumpulan data untuk penelitian ini yaitu dengan cara melakukan *Scraping* data *rating* pelanggan dan ulasan pelanggan menggunakan Webharvy.

Jumlah data yang harus dikumpulkan untuk membuat dataset ulasan pelanggan hotel Purwokerto di TripAdvisor.co.id, hanya mengambil 39 Hotel yang sudah memiliki ulasan dengan ulasan pelanggan sebanyak 1,166 ulasan. Data yang diperoleh selanjutnya dijadikan satu file lalu simpan dengan format .csv.

3.3.3 Pelabelan Data

Model analisis sentimen pada penelitian ini merupakan permasalahan *Supervised Learning*. *Supervised Learning* dibentuk dari dataset yang memiliki target atau label. Pada kolom komentar ulasan pelanggan hotel di TripAdvisor.co.id, selain menuliskan ulasan, pelanggan juga dapat memberikan *rating* pelayanan 10 sampai 50. Pelabelan data akan dilakukan dengan cara memilah ulasan data positif menjadi label Positif berdasarkan *rating* ulasan sebesar 40 dan 50 dan ulasan data negatif menjadi label Negatif berdasarkan *rating* ulasan sebesar 10, 20, dan 30.

3.3.4 Ekplorasi Data

Pada tahapan ini dilakukan untuk mempresentasikan data yang telah melalui tahapan pelabelan data dengan grafik atau bagan untuk menampilkan info yang tersembunyi didalam data. Pada penelitian ini akan digunakan pie chart untuk menampilkan total persentase yang harus mencapai 100%, dimana setiap potongan pie akan menampilkan ukuran presentase data positif dan negative

3.3.5 Text Pre-Processing

Text Pre-processing dilakukan untuk membentuk “Golden Dataset”. Pada penelitian ini, tahap *text pre-processing* dilakukan dengan 3 langkah, yaitu :

1. *Data Cleansing*

Data cleansing adalah membersihkan data dari noise seperti tanda baca dan karakter lain yang tidak penting. *Data cleansing* dimulai dengan *Case folding*. *Case folding* adalah proses penyamaan *case* dalam sebuah dokumen. Pada penelitian ini tahap pertama pada *pre-processing* dilakukan *case folding* untuk mempermudah pencarian. Tidak semua teks pada dataset konsisten dalam penggunaan huruf kapital. Oleh karena itu peran *case folding* dibutuhkan dalam mengkonversi keseluruhan teks dalam dataset menjadi suatu bentuk standar. Pada saat bersamaan *case folding* juga berfungsi untuk membuang karakter selain huruf. Dalam penelitian ini *case folding* mengkonversi teks menjadi huruf kecil atau *lowercase*.

Contoh :

Kalimat awal :

“Berikut ini adalah 5 negara dengan pendidikan terbaik di dunia adalah Korea Selatan, Jepang, Singapura, Hong Kong, dan Finlandia.”

Setelah dilakukan *case folding* :

“berikut ini adalah negara dengan pendidikan terbaik di dunia adalah korea selatan jepang singapura hong kong dan finlandia.”

2. *Stopword*

Stopword merupakan kata yang diabaikan dalam pemrosesan, kata-kata ini biasanya disimpan ke dalam *stop lists*. Karakteristik utama dalam pemilihan *stopword* biasanya adalah kata yang memiliki frekuensi kemunculan yang tinggi misalnya kata penghubung seperti “dan”, “atau”, “tapi”, “akan” dan lainnya. *Stopword* pada penelitian ini bertujuan untuk mengurangi jumlah kata pada dataset yang nantinya akan berpengaruh dalam kecepatan dan performa proses pemodelan.

Pada penelitian ini menggunakan library Sastrawi. Untuk memulai proses *stopword* harus menginstal library Sastrawi. Pada library ini *stopword* telah didefinisikan terlebih dahulu dan di simpan di *stop list*.

Contoh :

Kalimat awal :

“Andi kerap melakukan transaksi rutin secara daring atau online. Menurut Andi belanja online lebih praktis & murah.”

Setelah dilakukan *stopword* :

['andi', 'kerap', 'transaksi', 'rutin', 'daring', 'online', 'andi', 'belanja', 'online', 'praktis', 'murah']

3. *Tokenizing*

Tokenizing adalah proses pemotongan sebuah dokumen menjadi bagian-bagian, yang disebut dengan token (memisahkan kata, simbol, frase, dan entitas penting lainnya).

Pada penelitian ini *tokenizing* dilakukan menggunakan method vektorisasi dari library sklearn.

Contoh :

Kalimat awal :

“selama pandemi kuliah dilakukan rutin secara daring atau online.”

Setelah *tokenizing* :

['selama', 'pandemi', 'kuliah', 'dilakukan', 'rutin', 'secara', 'daring', 'atau', 'online']

3.3.6 Feature Selection

Algoritma yang digunakan dalam seleksi fitur yaitu *Term frequency inverse document frequency* (TF-IDF). Metode TF-IDF merupakan metode pembobotan term yang banyak digunakan sebagai metode pembandingan terhadap metode pembobotan baru. Pada metode ini, perhitungan bobot term t dalam sebuah dokumen dilakukan dengan mengalikan nilai *Term Frequency* dengan *Inverse Document Frequency*.

- Menghitung Term Frequency (TF), yakni frekuensi kemunculan term (t) pada dokumen (D). Contoh (terdapat kalimat) :

D1“Menginap seminggu disini karena ada tugas kantor, hotelnya instagrammable banget dan pelayanannya ramah. Pulang2 dikasih compliment gantungan ikonnya meotel dari Mbak yanti. Terimakasih, pasti lain kali balik nginep sini lagi”

D2“Saya nyaman menginap dihotel ini. Mulai dr kamar, kamar mandi dan fasilitas sangat bagus dan berfungsi dengan baik. Kmrn saya sempat dilayani oleh mas Riki ai siapa ituu, orangnya ramahh, senyum teruss, sopan bangettt. Itu menambah kesan yg baik bagi saya untuk hotel ini. Ga nyesel nginep disini pokoknya.”

D3“Hotel ini layak jadi pilihan pertama bila anda berkunjung ke kota Purwokerto. Lokasi yang strategis,kamar yang nyaman & kolam renangya jg aman (terdalam 1,2 m) menjadikan hotel ini layak jadi pilihan.”

Berikut penjelasanya pada Tabel 3.6.

Tabel 3.1 Contoh TF

Term(t)	D1(dokumen)	D2	D3
Pelayanan	1	0	0
Baik	0	2	0
...

- Menghitung Document Frequency (DF), yaitu banyaknya dokumen dimana suatu term (t) muncul. Contoh berdasarkan soal yang sama pada poin pertama ada pada Tabel 3.7.

Tabel 3.2 Contoh DF

Term(t)	DF
Pelayanan	1
Baik	2
...	...

- Menghitung Invers Document Frequency (IDF) ada pada Tabel 3.8.

Tabel 3.3 Contoh IDF

Term(t)	DF	IDF=log(D/DF)
Pelayanan	1	$\log(3/1)=0,477$
Baik	2	$\log(3/2)=0,176$
...

- Menghitung nilai TF-IDF ada pada tabel 3.9.

Tabel 3.4 Contoh TF-IDF

Term(t)	TF			IDF	TF-IDF=TF*IDF		
	D1	D2	D3		D1	D2	D3
Pelayanan	1	0	0	0,477	0,477	0	0
Baik	0	2	0	0,176	0	0,352	0
...

3.3.7 Model Klasifikasi

Membangun model klasifikasi untuk menentukan sebuah kalimat ulasan sebagai anggota class positif atau class negatif berdasarkan nilai perhitungan probabilitas dari rumus Random Forest(RF). Jika probabilitas kalimat tersebut untuk class positif lebih besar, maka kalimat tersebut termasuk ke dalam positif. Namun jika probabilitas untuk class positif lebih kecil dari class negatif, maka kalimat tersebut masuk ke dalam class negatif. Setelah dilakukan preprocessing maka sebuah input algoritma Random Forest diperoleh hasil TF-IDF dimana term yang dihasilkan merupakan atribut dari algoritma. Random forest sendiri merupakan algoritma ensemble learning yang mana untuk memperoleh keputusan akhir akan dilakukan votting majority.

Random forest pada konsepnya menggabungkan beberapa Decision Tree untuk melakukan klasifikasi, setiap tree akan diberikan data secara acak dari sebagian dataset training, metode yang digunakan untuk memperoleh sample dataset secara acak adalah metode bagging atau disebut juga Bootstrap Aggregating. Di mana data training akan di pilih secara acak dengan jumlah tertentu dan diletakan pada sebuah tempat penampung sampel data. Dari sampel data tersebut dilakukan

decision tree sejumlah yang ditentukan, hasil dari masing masing tree tersebut akan disimpan sebagai model yang digunakan saat uji coba.

Berikut contoh langkah-langkah dalam melakukan klasifikasi analisis sentimen :

Langkah-1: Membuat *bootstrapped* dataset.

Tabel 3.5 Contoh Original Dataset

no	Fitur 1	Fitur 2	Fitur 3	Fitur 4
1	0,3	0,3	0,2	0,3
2	0,5	0	0,3	0,4
3	0,3	0,3	0	0,2
4	0	0	0,3	0,3

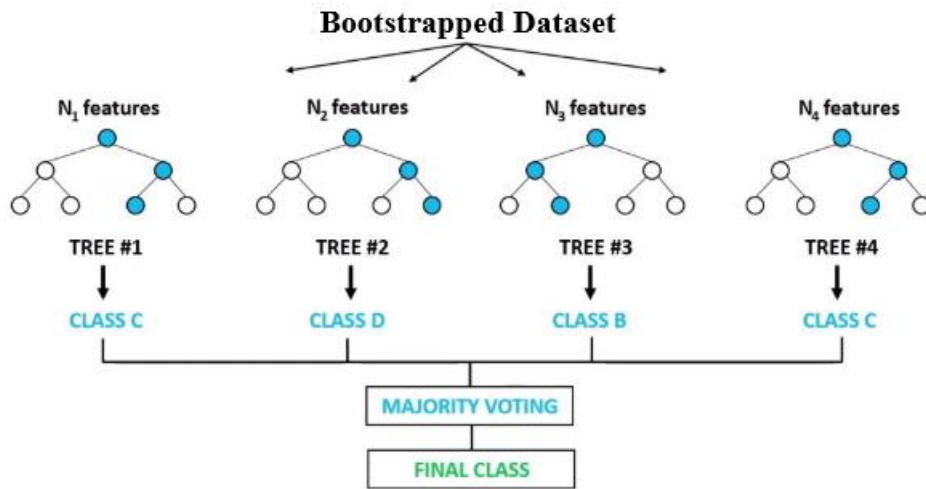
Tabel 3.6 Contoh Bootstrapped Dataset

no	Fitur 1	Fitur 2	Fitur 3	Fitur 4
1	0,5	0	0,3	0,4
2	0,3	0,3	0,2	0,3
3	0,5	0	0,3	0,4
4	0,3	0,3	0	0,2

Langkah-2: Buat pohon keputusan yang terkait dengan titik data yang dipilih (Subset). Kemudian akan didapatkan hasil prediksi dari setiap pohon keputusan yang dibuat.

Langkah-3: Pemungutan suara kemudian akan dilakukan untuk setiap hasil yang diprediksi.

Langkah-4: Terakhir, algoritme akan memilih hasil prediksi yang paling banyak dipilih sebagai prediksi akhir.



Gamabar 3.5 Contoh cara kerja Random Forest

3.3.8 Evaluasi dan Hasil

Untuk mengetahui akurasi, Confusion Matrix digunakan untuk mengetahui akurasi dari hasil algoritma *Random forest Classifier* dan *Term frequency inverse document frequency* (TF-IDF). Dengan menggunakan data *training* sebagai data validasi perhitungan untuk data *test*, maka akan menghasilkan nilai TP, TN, FP, FN sehingga akurasi dapat dihitung dan diketahui. Lalu apakah kedua algoritma tersebut dalam pemodelan analisis sentimen menggunakan data hotel di Purwokerto mendapatkan hasil akurasi yang baik atau tidak.