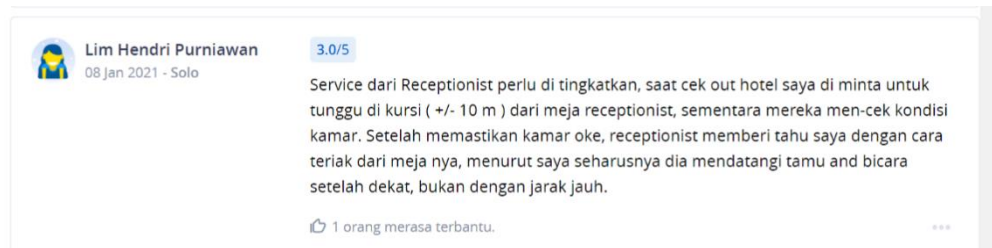


## BAB III METODOLOGI PENELITIAN

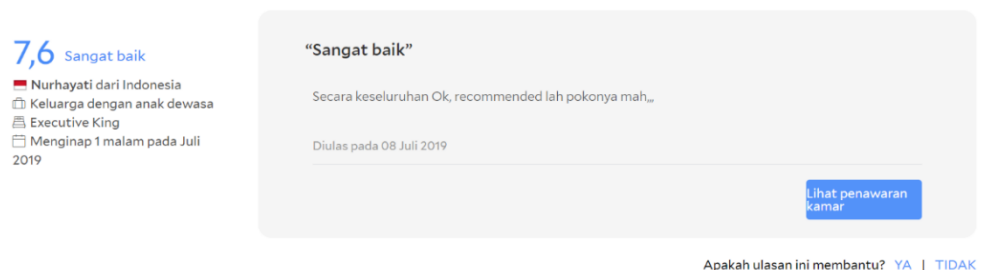
### 3.1 Subjek Dan Objek Penelitian

#### 3.1.1 Subjek Penelitian

Subjek pada penelitian merupakan pelanggan hotel Tasikmalaya yang sudah memberikan *review*-nya secara tertulis di Tiket sebanyak 2.185 ulasan dan Agoda sebanyak 1.110 ulasan. Berikut ini merupakan tampilan dan contoh ulasan di Tasikmalaya pada situs Tiket.com Gambar 3.1 dan Agoda pada Gambar 3.2



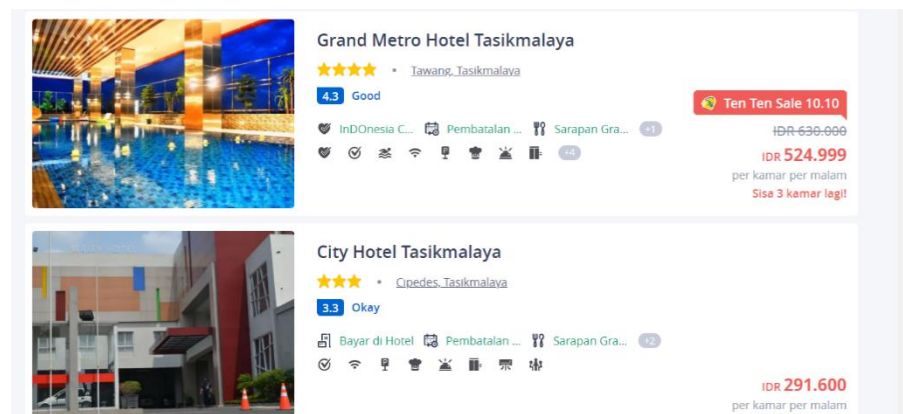
**Gambar 3.1 Contoh Ulasan Tiket.com**



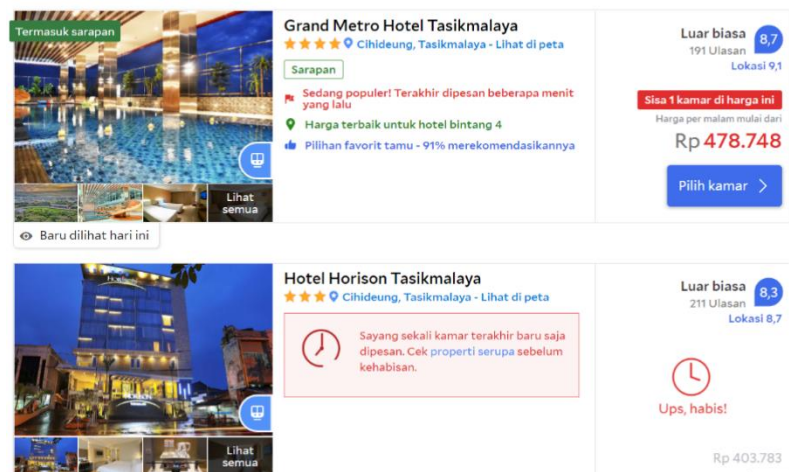
**Gambar 3.2 Contoh Ulasan Agoda**

#### 3.1.2 Objek Penelitian

Objek pada penelitian ini merupakan penginapan di Tasikmalaya yang sudah disediakan di Tiket dan Agoda. Tiket menyediakan 82 penginapan dan Agoda menyediakan 69 penginapan baik berbintang maupun tidak berbintang. Berikut contoh tampilan penginapan di Tiket pada Gambar 3.3. tampilan penginapan di Agoda pada Gambar 3.4



**Gambar 3.3 Contoh Tampilan Tiket.com**



**Gambar 3.4 Contoh Tampilan Agoda**

### 3.2 Alat Dan Bahan Penelitian

Alat yang digunakan dalam penelitian ini berupa perangkat keras (*hardware*) dan perangkat lunak pendukung (*software*). Bahan yang digunakan merupakan data ulasan pelanggan dari situs Tiket dan Agoda

#### 3.2.1 Alat Penelitian

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

1. *Manufacture* : ASUSTek Computer Inc.
2. Processor : Intel(R) Core(TM) i3-7100U CPU @ 2.40GHz 2.40 GHz
3. Installed memory / RAM : 12 GB (11,9 GB usable)

4. 64-bit operating system, x64-based processor

Adapun *hardware* dan *software* yang digunakan dalam penelitian sebagai berikut :

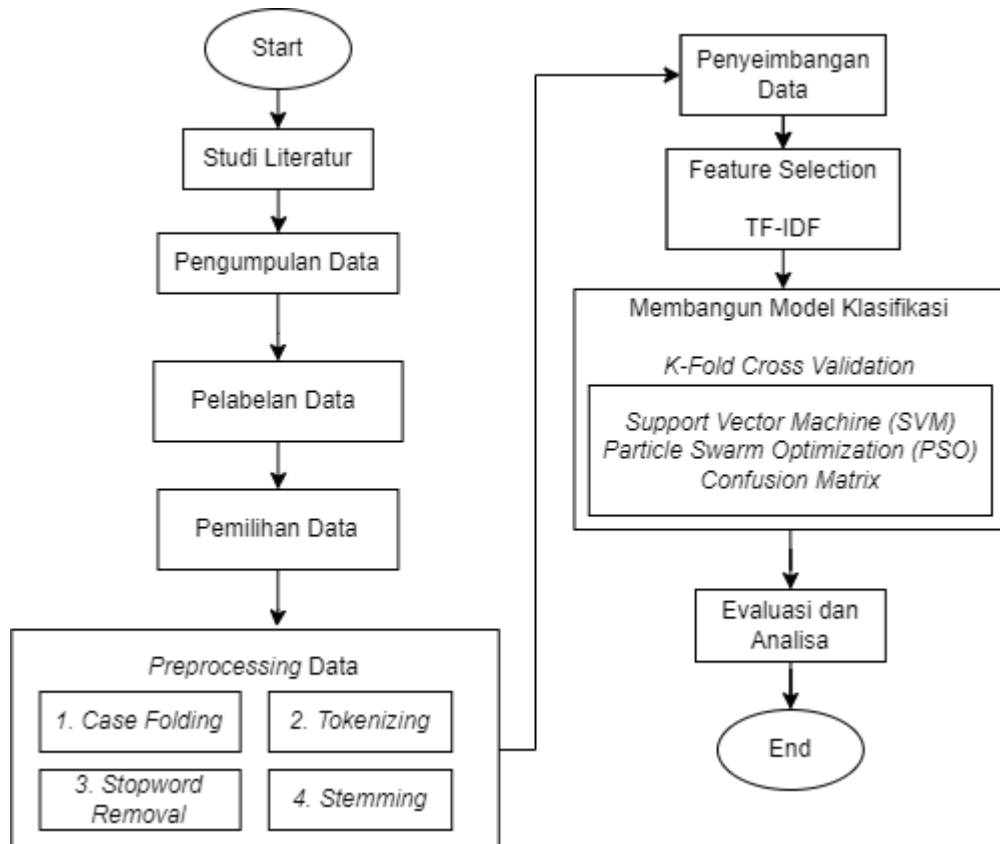
1. Sistem Operasi *Windows 10 Home Single Language*
2. *Microsoft excel*
3. *Data miner*
4. *Jupyter Notebook*

### 3.2.2 Bahan Penelitian

Bahan penelitian yang digunakan merupakan data ulasan pelanggan hotel-hotel Tasikmalaya yang diperoleh dari Tiket dan Agoda.

### 3.3 Diagram Alur Penelitian

Tahapan proses pada penelitian ditunjukkan oleh gambar. Tahapan proses penelitian terdiri dari :



**Gambar 3.5 Diagram Alur Penelitian**

### 3.3.1 Studi Literatur

Penelitian ini membutuhkan referensi sebagai dasar tahapan penelitian serta referensi yang digunakan penulis menggunakan jurnal terdahulu yang berkaitan dengan penelitian yang akan diambil sebagai referensi. Adapun buku terkait yang digunakan untuk menambah wawasan penulis dalam menyelesaikan penelitian ini.

### 3.3.2 Pengumpulan Data

Pengumpulan data untuk penelitian ini melalui cara *scraping* data ulasan pelanggan dan *rating* pelanggan hotel di Tasikmalaya pada Tiket.com dan Agoda menggunakan *Data Miner*. Jumlah data yang dikumpulkan untuk membuat dataset[32][41] :

1. Ulasan pelanggan hotel Tasikmalaya di Tiket.com dengan ulasan pelanggan sebanyak 2.185 ulasan dari 38 hotel.
2. Ulasan pelanggan hotel Tasikmalaya di Agoda dengan ulasan pelanggan sebanyak 1.110 ulasan dari 39 hotel.

### 3.3.3 Pelabelan Data

Model analisis sentimen pada penelitian menggunakan algoritma *Support Vector Machine* yang merupakan algoritma *supervised learning* dimana dataset memiliki label [32][41]. Pada kolom ulasan pelanggan hotel di Tiket terdapat terdapat ulasan komentar pelanggan serta rating pelayanan 1 sampai 5. Pelabelan data dilakukan dengan memilah ulasan positif menjadi 1 berdasarkan rating ulasan lebih dari sama dengan 3,0 dan ulasan negatif menjadi 0 berdasarkan rating ulasan kurang dari 3,0.

Agoda terdapat ulasan komentar pelanggan serta rating pelayanan 0 sampai 10. Pelabelan data dilakukan dengan memilah ulasan positif menjadi 1 berdasarkan rating ulasan lebih dari sama dengan 6,0 dan ulasan negatif menjadi 0 berdasarkan rating ulasan kurang dari 6,0.

### 3.3.4 Pemilihan Data

Pada tahap penelitian ini, dataset yang sudah dilakukan pelabelan data selanjutnya dilakukan pemilihan data dengan cara mengambil ulasan yang

memiliki karakter lebih dari 25 karakter. Karakter yang kurang dibawah 25 akan dihapus dan tidak akan digunakan untuk proses klasifikasi[32][41].

### 3.3.5 Preprocessing Data

Pada tahap *preprocessing* data, data diolah untuk klasifikasi teks terbagi menjadi beberapa tahapan. Tahapan yang digunakan untuk membantu menghasilkan nilai akurasi terbaik, yaitu :

#### 1. *Case Folding*

*Case Folding* merupakan tahap mengkonversi huruf kapital pada dokumen menjadi huruf standar yaitu *lowercase*[9][10][21][32].

Contoh :

##### a. Sebelum proses *case folding* :

“Tempatnya rapi, bersih, WiFi bagus, lokasi strategis”.

##### b. Setelah proses *case folding* :

“tempatny rapi, bersih, wifi bagus, lokasi strategis”.

#### 2. *Tokenizing*

*Tokenizing* merupakan proses pemecahan suatu kalimat menjadi per-kata serta penghapusan pada numerik, simbol, spasi (*whitespace*), dan tanda baca[9][10][21][32].

##### a. Sebelum *tokenizing* :

“kamar bersih, aman dan nyaman sekali serta harga terjangkau.”

##### b. Setelah *tokenizing* :

“kamar”,”bersih”,”aman”,”dan”,”nyaman”,”sekali”,”serta”,”harga”,”terjangkau”.

#### 3. *Stopword Removal*

Stopword merupakan pengabaian penghubung kata dan memiliki frekuensi kemunculan tinggi, seperti "dan", "atau", "tapi" "akan" dan lainnya. Untuk melaksanakan tahap ini dapat menggunakan library NLTK yang telah di definisikan terlebih dahulu dan di simpan di dalam stop list[9][10][21][32].

##### a. Sebelum *Stopword Removal* :

" Andi kerap melakukan transaksi rutin secara daring atau online.

Menurut Andi belanja online lebih praktis & murah "

b. Setelah *Stopword Removal* :

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

#### 4. *Stemming*

*Stemming* merupakan pengambilan kata dasar dalam suatu kata, dengan menghilangkan kata imbuhan atau *affixes* yang terdiri dari kata awal (*prefixes*), sisipan kata atau *infixes*, kata akhiran atau *suffixes* dan mengilangkan awalan kata dan akhiran kata atau *confixes*[9][10][21][32]. Library Sastrawi digunakan untuk melakukan *stemming*. Berikut merupakan langkah-langkah *stemming* :

- 1) Pemeriksaan langkah pertama ini akan menentukan langkah berikutnya. Apabila sudah berdasarkan kata dasar maka proses ini akan diperhentikan. Jika belum, maka proses akan dilanjutkan.
- 2) Kedua, menghilangkan kata *Inflection uffixes* (" -nya", "-mu", "-ku", "lah", dan "kah"). Jika kata termasuk dalam particles ("-lah", "-kah", "-tah" atau "-pun") berarti penghapusan ulang terdiri dari Possesive Pronouns ("-ku", "-mu", atau "-nya")
- 3) Menghilangkan *derivational suffix* atau imbuhan turunan dengan menghilangkan imbuhan ("-i", "-kan", "-an")
- 4) Menghilangkan *derivational prefix* atau imbuhan turunan, kata awalan seperti ("be-", "di-", "ke-", "me-", "pe-", "se-", dan "te-")
- 5) Jika sampai dalam tahap 4 tidak menemukan kata untuk dilakukan *stemming*, maka dilakukan analis apakah termasuk dalam kedalam tabel *diambiguitas* atau tidak.
- 6) Jika sampai langkah kelima tetap gagal maka, algoritma akan mengembalikan kata asli. Contoh :

7) Kata Awal :

"Anto sering melakukan pemesanan hotel secara online saat berlibur. Menurut Anto pemesanan hotel pinline lebih efektif & efisien".

8) Kata setelah di *Stemming* :

"Anto sering laku pesan hotel online saat libur  
turut anto pesan hotel lebih efektif efisien".

### 3.3.6 Penyeimbangan Data

Pada tahap penelitian ini, proses penyeimbangan data dilakukan dengan tujuan tidak adanya data yang tidak seimbang atau data lebih banyak ke ulasan positif daripada ke ulasan negatif maupun sebaliknya. Hal ini dilakukan dengan tujuan pada saat model klasifikasi data menjadi lebih seimbang baik ke ulasan positif maupun ke ulasan negatif sehingga data yang dihasilkan lebih baik[32][41]. Langkah yang dilakukan untuk menangani data yang tidak seimbang dapat dilakukan dengan metode berikut [32]:

1. *Undersampling*

Data diseimbangkan dengan mengurangi ukuran kelas data yang berlimpah. metode ini digunakan ketika jumlah data mencukupi. menjaga semua sampel di kelas langka dan secara acak memilih jumlah sampel yang sama di kelas berlimpah.

2. *Oversampling*

Digunakan pada saat jumlah data tidak mencukupi. Mencoba menyeimbangkan dataset dengan meningkatkan ukuran sampel langka. Daripada membuang sampel berlimpah, sampel langka baru dihasilkan dengan menggunakan SMOTE (Sintetis Minoritas Sampling Teknik).

### 3.3.7 Feature Selection Data

Pada tahap penelitian ini, proses *feature selection* data menggunakan TF-IDF dilakukan dengan tujuan meningkatkan akurasi klasifikasi dengan mengubah data berupa kata( $f$ ) menjadi numerik. Proses dimulai dengan menganalisa sebuah kata pada dataset( $d$ ) dan menghitung nilai kemunculan kata yang diperoleh dari dataset yang disebut *TF*. Tahap berikutnya menghitung banyaknya dataset dari kata yang muncul kemudian menghitung kata yang di distribusikan pada dataset sehingga menghasilkan nilai atau bisa

disebut dengan *IDF*. Setelah itu kedua nilai tadi dikalikan akan menghasilkan sebuah kata dengan nilai yang terbesar hingga yang terkecil untuk dapat mengetahui seberapa pentingnya dalam fitur tersebut pada sebuah penelitian ini[32].

### 3.3.8 Membangun Model Klasifikasi

Pada tahap penelitian ini, data yang sudah diseleksi kemudian diproses menggunakan model klasifikasi dengan metode *Support Vector Machine* dan *Particle Swarm Optimization* yang akan diimplementasikan menggunakan bahasa pemrograman Python. Model klasifikasi penelitian ini terdapat proses *K-Fold Cross Validation*. *K-Fold Cross Validation* melakukan pengacakan pada dataset yang sudah diolah menjadi beberapa dataset[10][21][42]. Pada penelitian *K-Fold Cross Validation* yang akan digunakan yaitu  $k=5$  dan  $k=10$ .

Pada penelitian pertama menggunakan *K-Fold Cross Validation* menggunakan  $k=5$  yang artinya dataset diuji secara acak sebanyak lima kali pengulangan dengan parameter dataset dibagi menjadi data *training* sebesar 80% yang akan digunakan sebagai proses melatih mesin/program tentang pengetahuan dataset penelitian dan data *test* sebesar 20% digunakan sebagai pengujian penelitian untuk mengetahui nilai akurasi dan klasifikasi dari penelitian. Pada pemrosesan *K-Fold Cross Validation* melakukan pengklasifikasian *Support Vector Machine* menggunakan *model kernel linear*, *kernel RBF (Radial Basis Function)*, dan *kernel polynomial*. Dataset yang sudah diproses lalu diuji dengan menggunakan algoritma *Support Vector Machine* terlebih dahulu kemudian diuji menggunakan *K-Fold Cross Validation* sehingga didapatkan hasil *Confusion Matrix* dari pemrosesan algoritma *Support Vector Machine*. Penelitian selanjutnya pada pemrosesan *K-Fold Cross Validation* melakukan pengklasifikasian *Support Vector Machine* dan *Particle Swarm Optimization* menggunakan *model kernel linear*, *kernel RBF (Radial Basis Function)*, dan *kernel polynomial*. Dataset yang sudah diproses lalu diuji dengan menggunakan algoritma *Support Vector Machine* dan *Particle Swarm Optimization* kemudian diuji menggunakan *K-Fold Cross*



*Validation* sehingga didapatkan hasil *Confusion Matrix* dari pemrosesan algoritma *Support Vector Machine* dan *Particle Swarm Optimization*.

Pada penelitian kedua menggunakan *K-Fold Cross Validation* menggunakan  $k=10$  yang artinya dataset diuji secara acak sebanyak sepuluh kali pengulangan dengan parameter dataset dibagi menjadi data *training* sebesar 90% yang akan digunakan sebagai proses melatih mesin/program tentang pengetahuan dataset penelitian dan data *test* sebesar 10% digunakan sebagai pengujian penelitian untuk mengetahui nilai akurasi dan klasifikasi dari penelitian. Pada pemrosesan *K-Fold Cross Validation* melakukan pengklasifikasian *Support Vector Machine* menggunakan *model kernel linear*, *kernel RBF (Radial Basis Function)*, dan *kernel polynomial*. Dataset yang sudah diproses lalu diuji dengan menggunakan algoritma *Support Vector Machine* terlebih dahulu kemudian diuji menggunakan *K-Fold Cross Validation* sehingga didapatkan hasil *Confusion Matrix* dari pemrosesan algoritma *Support Vector Machine*. Penelitian selanjutnya pada pemrosesan *K-Fold Cross Validation* melakukan pengklasifikasian *Support Vector Machine* dan *Particle Swarm Optimization* menggunakan *model kernel linear*, *kernel RBF (Radial Basis Function)*, dan *kernel polynomial*. Dataset yang sudah diproses lalu diuji dengan menggunakan algoritma *Support Vector Machine* dan *Particle Swarm Optimization* kemudian diuji menggunakan *K-Fold Cross Validation* sehingga didapatkan hasil *Confusion Matrix* dari pemrosesan algoritma *Support Vector Machine* dan *Particle Swarm Optimization*.

### **3.3.9 Evaluasi dan Analisa**

Penelitian menggunakan *Support Vector Machine* dan *Particle Swarm Optimization* untuk menganalisis sentimen terhadap ulasan hotel di Tiket.com dan Agoda khususnya di Tasikmalaya. Pengujian *K-Fold Cross Validation* dengan nilai  $k=5$  dan  $k=10$ . Algoritma *Support Vector Machine* sebagai algoritma utama untuk menganalisis sentimen dan *Particle Swarm Optimization* digunakan untuk meningkatkan akurasi. Adapun kernel yang digunakan yaitu *kernel linear*, *kernel Radial Basis Function (RBF)* dan *kernel Polynomial*. *Confusion matrix* digunakan untuk mendapatkan nilai *precision*,

*recall*, dan *accuracy* dari hasil model klasifikasi *Support Vector Machine* dan *Particle Swarm Optimization*. Setelah mendapatkan hasil evaluasi dari penelitian ini selanjutnya dapat ditarik kesimpulan dari penelitian yang sudah dilakukan [9].