

**TUGAS**  
**IF4072 PEMROSESAN BAHASA ALAMI**  
**Analisis Sentimen Berbasis Aspek pada *Review* Hotel**



**Oleh:**  
13515035 - Oktavianus Handika  
13515075 - Adrian Mulyana Nugraha

## Deskripsi aplikasi

Aplikasi mengklasifikasikan *review* pada suatu hotel. Klasifikasi tidak hanya dilakukan secara keseluruhan *review*, namun diklasifikasikan berdasarkan aspek-aspek yang ada pada *review* tersebut. *Review* diklasifikasi berdasarkan beberapa aspek pada umumnya seperti kualitas layanan, lokasi hotel, fasilitas hotel, dll.

## Latar belakang aplikasi

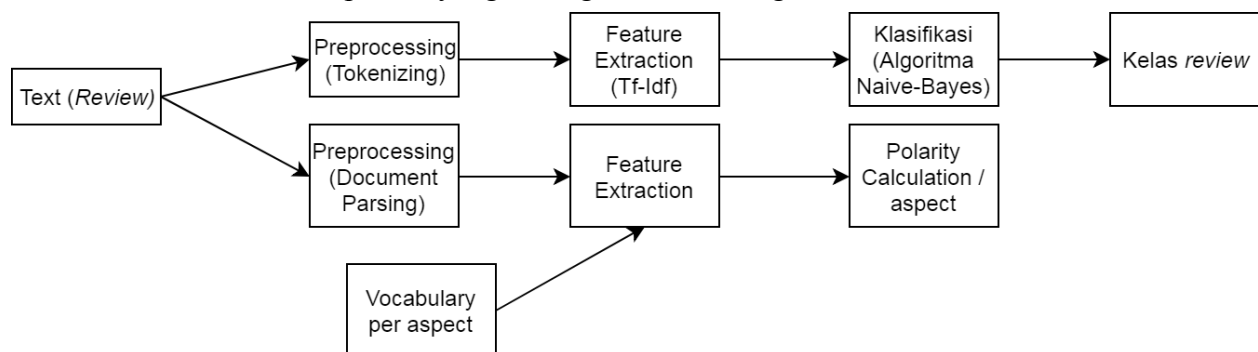
Terkadang *review* yang tersedia pada sebuah aplikasi mengenai suatu tempat atau objek wisata sudah terlalu banyak, tidak terstandar dan dapat membingungkan pembaca yang ingin mencari *review* mengenai tempat tersebut. Dengan adanya aplikasi yang dapat merangkum dan mengklasifikasi seluruh *review*, maka kualitas dari tempat tersebut dapat ditampilkan dari berbagai aspek tanpa membaca seluruh *review*. Selain itu, dapat membantu memberi gambaran mengenai hotel yang akan ditempati berdasarkan rangkuman *review* tersebut.

## Aplikasi lain yang sudah ada

Terdapat beberapa *project* yang telah membangun aplikasi untuk menentukan sentimen dari aspek-aspek pada suatu *review* atau komentar. Beberapa *project* yang telah dibuat untuk klasifikasi *review* berbasis aspek seperti proposal pada paper “Aspect based Sentiment Oriented Summarization of Hotel Reviews” (<https://www.sciencedirect.com/science/article/pii/S1877050917319439>) dan Aplikasi oleh JituShinde77 ([https://github.com/JituShinde77/ABSA\\_Project](https://github.com/JituShinde77/ABSA_Project))

## Arsitektur aplikasi

Berikut adalah arsitektur aplikasi yang dibangun untuk mengklasifikasikan *review*



## Teknik yang digunakan

Dalam aspect-based text classification yang kami buat ini, pertama dokumen di parse dan tokenize, kemudian dengan menggunakan Tf-Idf mengubah tiap kata menjadi bag of words. Kemudian algoritma Naive-Bayes akan mengelompokkan kemunculan kata yang sesuai aspek menjadi kelompok aspek positif atau negatif.

## Eksperimen (informasi data, skenario eksperimen, hasil eksperimen dan analisis hasil eksperimen)

Pada aplikasi ini, eksperimen dilakukan menggunakan data *review* hotel pada web <https://www.kaggle.com/anu0012/hotel-review> yang mana merupakan kumpulan *review* suatu hotel dari berbagai aplikasi yang telah disajikan menjadi dataset yang siap digunakan untuk pemrosesan *text*. Untuk eksperimen, digunakan 250 data *training* dan 250 data *test* dari sekitar 32 ribu data yang ada pada berkas untuk mempercepat proses eksperimen dan pengolahan data pada dataset.

Untuk klasifikasi *text review* berdasarkan sentimen secara keseluruhan, dilakukan *training* dataset dengan melakukan *preprocessing* dataset dengan tokenisasi pada dataset. Setelah dilakukan *preprocessing*, dilakukan ekstraksi fitur dengan menggunakan Tf-Idf. Kemudian, dataset akan dilatih dengan suatu algoritma untuk menentukan secara keseluruhan suatu *review* merupakan *review* dari pengguna yang menunjukkan bahwa pengguna senang terhadap suatu hotel atau tidak. Setelah itu, barulah model dilatih dengan algoritma *machine learning* untuk melatih data *text* hasil *review*-nya.

Untuk klasifikasi *text* ini, didapat hasil sebagai berikut:

*Accuracy score*: 66.8%

	Precision	Recall	F1-score
“happy”	0.66	1	0.8
“Not happy”	1	0.06	0.11
Avg / total	0.78	0.67	0.55

Berdasarkan hasil eksperimen yang diperoleh, dapat dilihat bahwa seluruh kesalahan / *error* yang terjadi pada *classifier* adalah kesalahan dalam menentukan *review* yang seharusnya memiliki sentimen ‘*not happy*’, namun *classifier* selalu memprediksi bahwa *review* tersebut termasuk dalam kategori ‘*happy*’. *Classifier* masih sulit mencari pembeda antara *review* yang memiliki sentimen ‘*happy*’ dengan ‘*not happy*’.

Pada analisis sentimen suatu aspek pada *review*, dilakukan pendekatan *rule-based* untuk menentukan apakah suatu aspek memiliki sentimen positif, negatif, atau netral. Suatu *review* akan di-*parse* menjadi beberapa kalimat. Setiap kalimat ini akan ditentukan apakah kalimat mengacu pada suatu aspek atau tidak sama sekali. Suatu kalimat dapat mengacu pada lebih dari 1 aspek. Nilai sentimen suatu kalimat pada suatu aspek ditentukan menggunakan ekstensi yang sudah ada, yaitu TextBlob. Nilai sentimen suatu kalimat dengan kalimat lainnya pada suatu *review* akan dikalkulasi sehingga membentuk sentimen per aspek suatu *review*.

Dari eksperimen menggunakan pendekatan ini, masih belum cukup akurat. Terdapat beberapa sentimen yang masih kurang sesuai seperti berikut:

1. Suatu sentimen seharusnya memiliki polaritas karena memiliki kata kunci terkait suatu aspek justru aspek tersebut masih bernilai netral.
2. Polaritas sentimen suatu kalimat tidak sesuai karena terdapat 2 kata dengan polaritas sentimen berbeda namun kekuatannya berbeda dalam suatu kalimat.
3. Kalkulasi *score* per aspek atau penentuan *threshold* suatu aspek termasuk positif atau negatif yang masih belum sesuai.

## **Pembagian Tugas**

Oktavianus Handika (13515035): Implementasi kode dan rule pada aspect sentiment analysis, dokumen dan pengujian

Adrian Mulyana Nugraha (13515075): Pembuatan classifier, dokumen