



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB II

TINJAUAN PUSTAKA

2.1 Sentiment Analysis

Menurut Bing (2012) *sentiment analysis*, atau disebut juga *opinion mining*, adalah bidang studi yang menganalisa opini, perasaan, evaluasi, penilaian, perilaku, dan emosi seseorang terhadap entitas seperti produk, servis, organisasi, individu, isu, acara, dan atribut mereka.

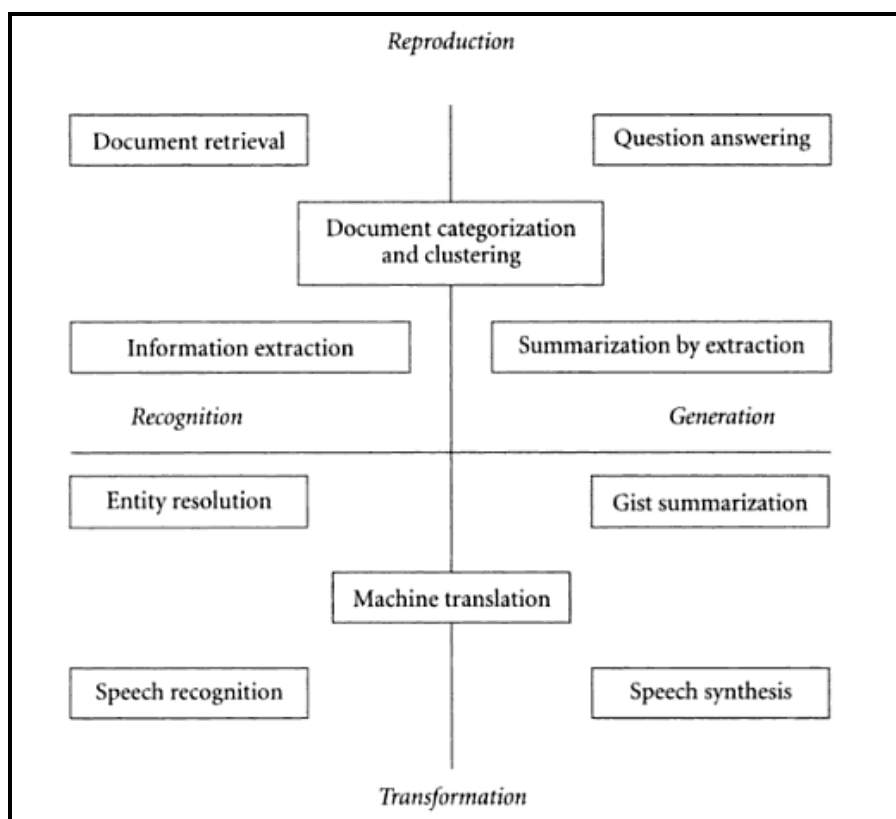
Bing (2012) juga mengatakan *sentiment analysis* adalah permasalahan *natural language processing*. Dia menyentuh semua aspek dari *natural language processing* seperti *coreference resolution*, *negation handling*, dan *word sense disambiguation*. Selain itu, *sentiment analysis* adalah permasalahan *natural language processing* yang sangat terbatas karena sistem tidak perlu sepenuhnya mengerti *semantics* dari setiap kalimat atau dokumen tetapi hanya perlu mengerti beberapa aspek dari hal tersebut, sebagai contoh, penilaian positif dan negatif dan target entitas atau topik.

2.2 Natural Language

Ela (2011) berpendapat bahwa *natural language* adalah bahasa yang secara natural berevolusi dan digunakan manusia dengan tujuan berkomunikasi, sebagai contoh, bahasa Inggris, bahasa Perancis, bahasa Jerman adalah *natural language*. *Natural language processing* atau NLP (dapat juga disebut *Computational Linguistics*) adalah studi ilmiah mengenai bahasa dilihat dari perspektif komputasi.

Dalam buku yang ditulis oleh Jackson, dikatakan bahwa ada permasalahan mendasar yang tidak dapat dihindari dan perlu diperhatikan yaitu :

1. Bagaimana sistem pencarian dapat memuaskan kebutuhan pengguna.
2. Kapan sebuah program dapat diandalkan dalam men ekstrak hal penting dari sebuah teks bebas.
3. Bagaimana cara mengevaluasi sebuah program text classification secara otomatis.
4. Apa yang membuat sebuah ringkasan yang baik dari sebuah dokumen.



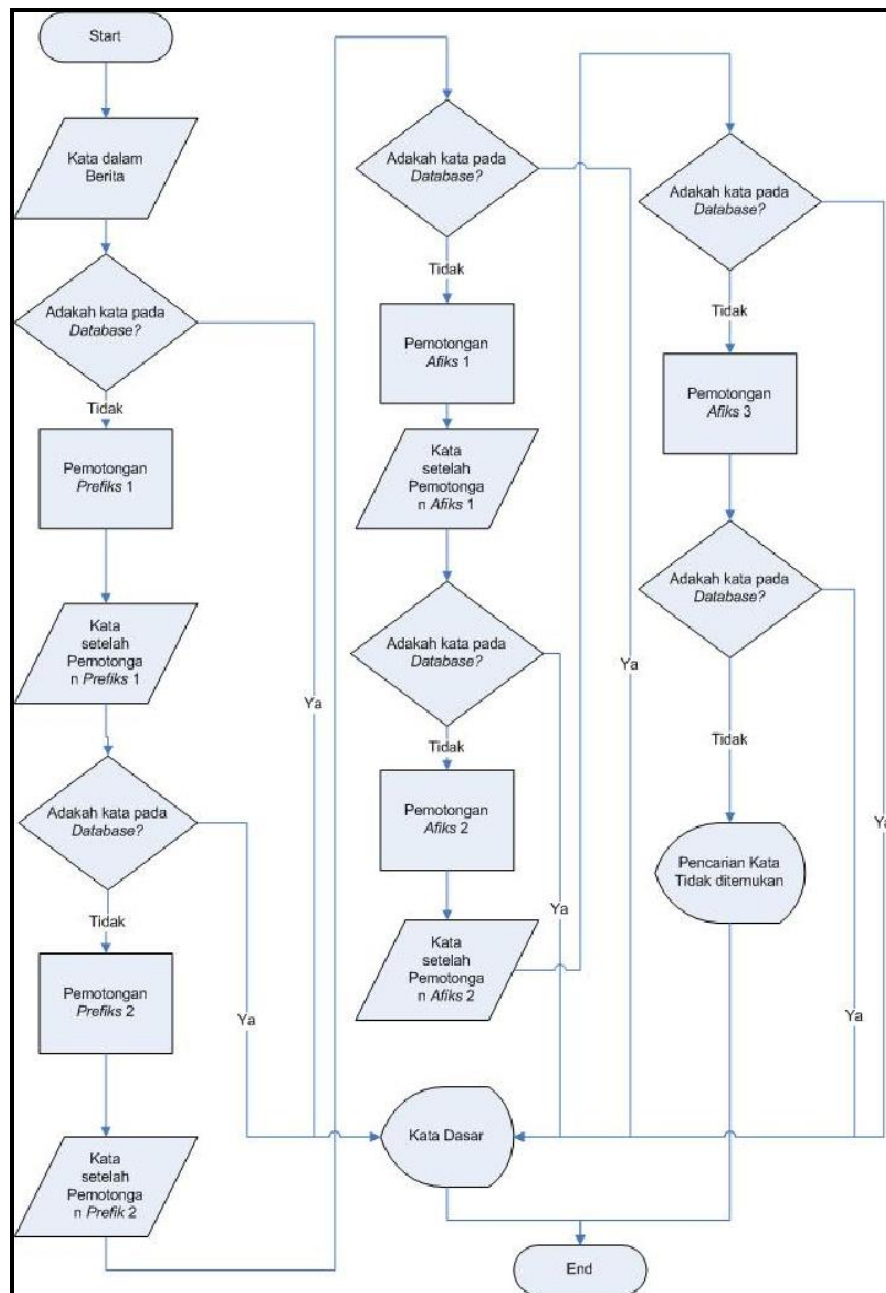
Gambar 2.1 Aplikasi pada Natural Language
Sumber : Jackson (2007)

2.3 Stemming dan Nazief Adriani

Liyantanto (2011) mengatakan bahwa *stemming* adalah salah satu cara yang digunakan untuk meningkatkan performa *information retrieval* dengan cara mentransformasi kata-kata dalam sebuah dokumen teks ke bentuk kata dasarnya.

Selain itu menurut Liyantanto (2011) tidak banyak algoritma yang dikhususkan untuk *stemming* bahasa Indonesia dengan berbagai keterbatasan di dalamnya. Salah satunya adalah algoritma Nazief & Adriani sebagai algoritma *stemming* untuk teks berbahasa Indonesia yang memiliki kemampuan prosentase keakuratan yang lebih baik dari algoritma lainnya.

Algoritma Nazief & Adriani dibuat oleh Bobby Nazief dan Mirna Adriani memiliki tahap-tahap seperti dijelaskan pada gambar 2.2 berikut.



Gambar 2.2 Diagram alir Nazief & Adriani
Sumber : Liyantanto (2011)

2.4 Text Classification

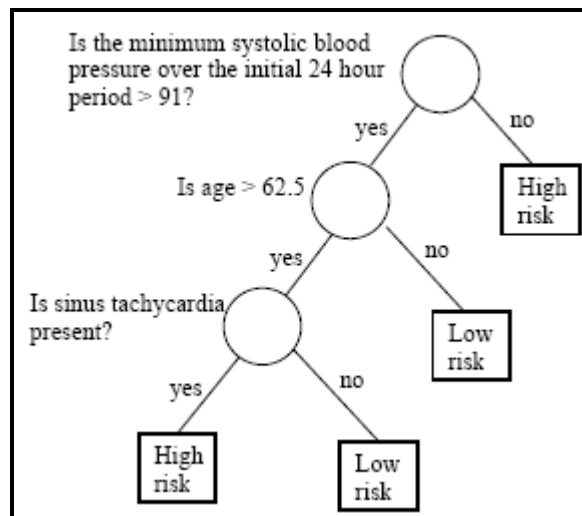
Menurut Nedjah (2009), text classification adalah sebuah pekerjaan penting yang digunakan pada banyak aplikasi. Hal ini dapat dilakukan dengan memberikan label kepada dokumen berdasarkan kumpulan dokumen

sebelumnya yang sudah diberi label. Langkah pertama adalah merepresentasikan dokumen sebagai vektor dalam sebuah *high-dimension vector*, dimana masing-masing dimensi menyesuaikan kepada nilai sebuah fitur, atau secara spesifik, sebuah *term*. Program kemudian menggunakan vektor tersebut untuk mengklasifikasikan dokumen ke kategori yang tepat. Untuk melakukan hal itu sangatlah penting untuk menentukan relevansi setiap *term*, atau dikenal sebagai *term weighting*.

2.5 Decision Trees

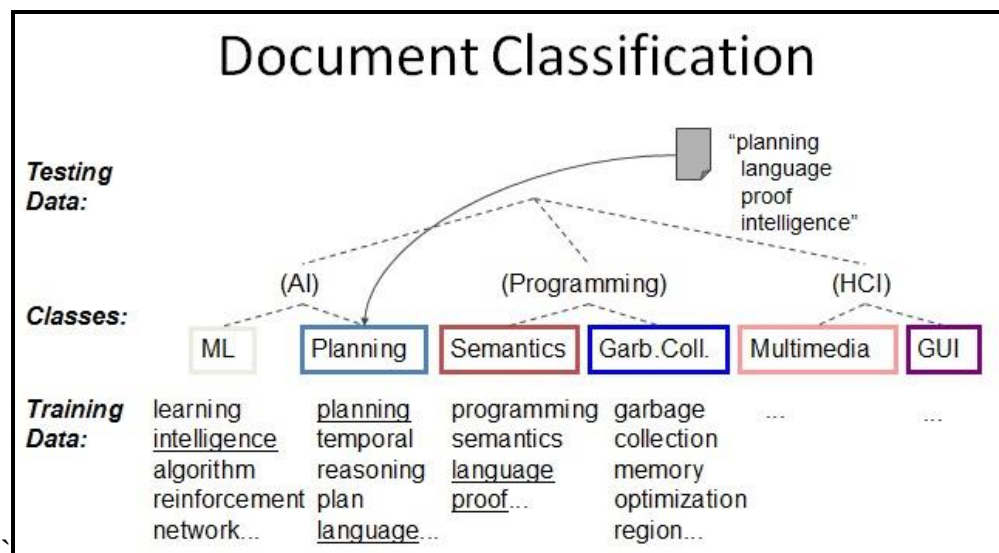
Rokach (2008) menyatakan bahwa *decision tree* adalah diagram analisis yang dapat membantu mengambil keputusan ketika memilih diantara opsi berbeda dengan memproyeksikan hasil yang mungkin. *Decision tree* memberikan gambaran dari beberapa tahapan yang akan memberikan setiap kemungkinan hasil.

Menurut Rokach (2008) dalam *data mining*, *decision tree* adalah model prediktif yang dapat digunakan untuk pengklasifikasian. Ketika *decision tree* digunakan untuk tugas pengklasifikasian, akan lebih cocok disebut *classification tree*. *Classification tree* digunakan untuk mengklasifikasikan sebuah obyek atau kejadian kepada sekumpulan kelas yang sudah didefinisikan berdasarkan nilai atributnya.



Gambar 2.3 Classification Tree dalam dunia medis
sumber : <https://onlinecourses.science.psu.edu/stat857/node/22>

Ketika membangun *text classification* dengan *decision tree* maka tahap pertama adalah membangun *decision tree* terpisah bagi setiap kategori. Setelah melakukan perhitungan maka akan ditentukan melalui jumlah kata yang muncul di setiap kategori.



Gambar 2.4 Klasifikasi dokumen menggunakan Decision Tree
sumber : <http://clic.cimec.unitn.it/massimo/Teach/AI>

Gambar 2.4 menunjukkan penggunaan *classification tree* pada pengklasifikasian dokumen. Dokumen yang masuk dapat diklasifikasikan dengan melihat konten dari dokumen tersebut. Konten dibandingkan dengan *training data* yang sudah ada pada setiap kelasnya dan dihitung jumlah terbanyak *training data* dalam setiap kelas. Dokumen diklasifikasikan sebagai kelas yang memiliki jumlah kesamaan antar *training data* dengan testing data.

2.6 Social Media

Dalam artikel yang ditulis oleh Tea (2014), media sosial adalah saluran atau sarana pergaulan sosial secara online di dunia maya. Para pengguna media sosial berkomunikasi, berinteraksi, saling kirim pesan, dan saling berbagi. Media sosial yang paling populer digunakan di Indonesia antara lain adalah Twitter. Sebagai salah satu media komunikasi, media sosial tidak hanya dimanfaatkan untuk berbagi informasi dan inspirasi, tapi juga ekspresi diri dan ajang bercerita bahkan berkeluh kesah maupun sumpah serapah.

2.7 Twitter

Pada tahun 2006, sebuah layanan komunikasi bernama Twitter diperkenalkan. Awalnya Twitter hanya sebuah poyek sampingan dari perusahaan *podcasting* di San Francisco, namun tidak lama kemudian proyek sampingan ini telah menjadi pertunjukan utama. Twitter telah menjadi sebuah kunci komunikasi saat terjadi acara politik dan bencana alam. Selain itu bisnis saat ini bergantung pada Twitter untuk *marketing*, *public relation* dan *customer service*.

O'Reilly (2011) mengatakan Twitter adalah sebuah layanan pesan yang berbagi banyak karakteristik dengan alat komunikasi yang selama ini digunakan. Twitter menjadi unik karena pesan yang diterima dan dikirim hanya berjumlah 140 karakter atau hanya sepanjang tajuk utama berita. Hal ini menjadikan pesan tersebut mudah ditulis dan dibaca. Selain itu pesan dalam Twitter bersifat publik dan tidak mengharuskan memberi izin kepada orang lain untuk melihat apa yang kita tulis.

2.8 Opini

Menurut Imadha (2013), opini adalah pendapat, ide, atau pikiran untuk menjelaskan kecenderungan atau preferensi tertentu terhadap perspektif dan ideologi akan tetapi bersifat tidak objektif karena belum mendapatkan pemastian atau pengujian. Opini bukanlah merupakan sebuah fakta, akan tetapi jika di kemudian hari dapat dibuktikan atau diverifikasi, maka opini akan berubah menjadi sebuah kenyataan atau fakta.

2.9 Uji Data

Zegarelli (2007) mengatakan untuk dapat membuat data statistik membuat persen dapat dilakukan dengan cara sebagai menulis sebuah pernyataan yang berisi jumlah anggota dan jumlah total dalam sample. Jumlah persentase dapat ditulis dengan rumus 2.1 dibawah ini.

$$\frac{\text{number who share attributes}}{\text{number in sample}} \times 100\% \dots\dots\dots \text{Rumus 2.1}$$