

BAB 2

LANDASAN TEORI

2.1 Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen[8].

Analisis Sentimen, atau yang dikenal juga dengan Opinion Mining adalah ilmu yang menganalisis opini seseorang, sentimen, perilaku, penilaian dan emosi terhadap suatu entitas, yaitu seperti produk, pelayanan, organisasi, dan individu. Terdapat beberapa jenis lainnya yang memiliki tugas yang sedikit berbeda seperti opinion extraction, sentimen mining, subjectivity analysis, emotion analysis, review mining. Secara umum, analisis sentimen dibagi menjadi tiga level bagian utama, yaitu level dokumen, level kalimat, dan level aspek dan entitas[3].

2.2 Aspek

Aspek merupakan pemunculan suatu gagasan, masalah, peristiwa atau situasi sebagai pertimbangan yang dilihat dari sudut pandang tertentu, dilihat dari segi waktu yang menyertai keberlangsungan situasi tersebut. Istilah aspek (terjemahan dari bahasa Inggris *aspect*) itu sebenarnya merupakan terjemahan dari Bahasa Rusia *vid*. Adapun *vid* dalam Bahasa Rusia(Slavia) secara konseptual adalah sebagai berikut:

“[...] orang boleh berbicara tentang aspek hanya apabila makna-makna aspektualitas tertentu dari sebagian besar (kadang-kadang bahkan

keseluruhan) leksikon verba diungkapkan secara tetap melalui oposisi bentuk-bentuk gramatikal paradigmatik dari verba yang sama”

Konsep Aspek tersebut dapat mengandung arti bahwa suatu gejala aspektualitas dapat disebut aspek hanya apabila memenuhi persyaratan berikut:

1. Makna aspektualitas diungkapkan melalui proses morfologi yang teratur dan meliputi keseluruhan atas sebagian terbesar leksikon verba.
2. Proses morfologi itu bersifat paradigmatik.

Berdasarkan konsep itulah maka aspek termasuk dalam kategori gramatikal[2].

2.3 Bayesian Network

Bayesian Network (BN), juga dikenal sebagai jaringan kepercayaan (atau Bayes nets untuk pendek), probabilistik *graphical models*(GM). Ini struktur grafis digunakan untuk merepresentasikan pengetahuan tentang domain yang tidak pasti. Secara khusus, setiap node dalam grafik mewakili variabel acak, sementara ujung-ujungnya antara node mewakili probabilistik dependensi antara variabel acak yang sesuai. Dependensi kondisional dalam grafik sering diestimasi dengan menggunakan statistik dan metode perhitungan. Oleh karena itu, BN menggabungkan prinsip dari teori grafik, teori Probabilitas, komputer sains, dan statistik.

BN sesuai dengan struktur GM lain yang dikenal sebagai grafik asiklik diarahkan (DAG) yang populer di statistik, pembelajaran mesin, dan buatan masyarakat intelijen. BN keduanya secara matematis teliti dan intuitif dimengerti. Mereka memungkinkan representasi yang efektif dan perhitungan *joint probability distribution*(JPD) lebih dari satu set variabel acak[9].

Struktur DAG didefinisikan oleh dua set: set node (simpul) dan set tepi yang diarahkan. Simpul mewakili variabel acak dan Digambar sebagai lingkaran yang diberi label oleh nama variabel. Ujung-ujungnya mewakili ketergantungan langsung di antara variabel dan digambar dengan tanda panah di antara simpul. Khususnya, sebuah tepi dari simpul X_i ke simpul X_j merupakan statistic

ketergantungan antara variabel yang sesuai. Dengan demikian, tanda panah menunjukkan bahwa nilai diambil oleh variabel X_j tergantung pada nilai yang diambil oleh variabel X_i , atau berbicara kasar bahwa variabel X_i “memengaruhi” X_j . Node X_i kemudian disebut sebagai parents dari X_j dan, sama, X_j disebut sebagai anak dari X_i . Perpanjangan dari istilah-istilah silsilah ini sering digunakan untuk menentukan himpunan "keturunan" - set node yang bisa dihubungi secara langsung path dari node, atau node “leluhur” – set node dari mana node dapat dijangkau pada jalur langsung. Struktur asiklik grafik menjamin bahwa tidak ada simpul yang bisa leluhurnya sendiri atau keturunannya sendiri. Misalnya kondisi sangat penting untuk faktorisasi dari Probabilitas gabungan dari kumpulan node sebagai lihat di bawah ini. Perhatikan bahwa meskipun panah mewakili hubungan kausal langsung antara variabel, yang proses penalaran dapat beroperasi pada BNs dengan menyebarkan informasi ke segala arah.

BN mencerminkan kemandirian bersyarat yang sederhanapernyataan. Yakni setiap variabel bersifat independent dari nondescendents-nya dalam grafik yang diberikan negara dari parentsnya. Properti ini digunakan untuk mengurangi, kadang-kadang secara signifikan, jumlah parameter yang diperlukan untuk mengkarakterisasi JPD dari variabel. Pengurangan ini memberikan cara yang efisien untuk menghitung Probabilitas posterior yang diberikan bukti.

Selain struktur DAG, yang sering dianggap sebagai bagian "kualitatif" dari model, satu perlu menentukan parameter "kuantitatif" dari model. Parameter dijelaskan dengan cara yang konsisten dengan properti Markovian, di mana distribusi Probabilitas bersyarat (CPD) di masing-masing simpul hanya bergantung pada parentsnya. Untuk diskrit acak variabel, Probabilitas kondisional ini sering diwakili oleh sebuah tabel, daftar kemungkinan lokal bahwa simpul anak mengambil setiap nilai yang layak – untuk setiap kombinasi nilai dari parent[9]. Sendi distribusi kumpulan variabel dapat ditentukan secara unik oleh Probabilitas kondisional lokal ini tabel (CPT).

Dalam pemodelan Probabilitas, misalnya data sering dianggap sebagai peristiwa, pengamatan, atau realisasi dari variabel acak yang mendasarinya. Diberikan variabel acak diskrit A , $P(A)$ adalah fungsi yang mengkodekan

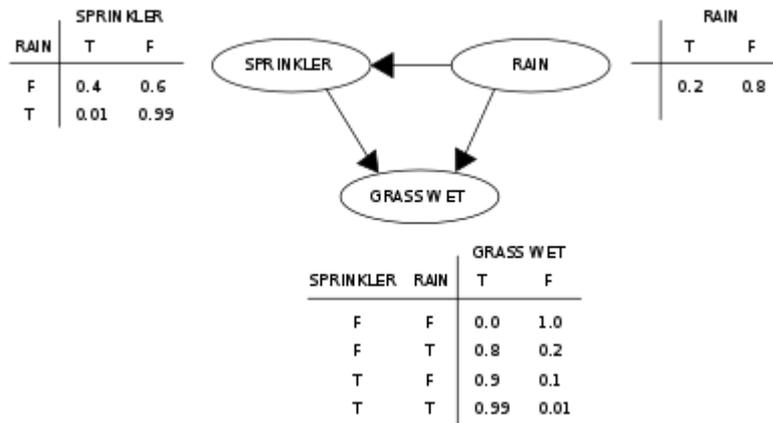
Probabilitas untuk masing-masing kategori, kelas, atau menyatakan bahwa A banyak berada. Demikian pula, untuk variabel acak kontinu x , $p(x)$ adalah fungsi yang menetapkan kepadatan Probabilitas untuk semua kemungkinan nilai x . Sebaliknya, $P(A = a)$ adalah Probabilitas tunggal untuk mengamati peristiwa spesifik $A = a$. Notasi ini sering disederhanakan menjadi $P = a$, tetapi orang perlu berhati-hati untuk mengingat apakah a didefinisikan sebagai variabel acak atau sebagai pengamatan. Demikian pula untuk pengamatan bahwa variabel acak kontinu x memiliki nilai x_1 atau nilai x_1 : itu umum untuk tulis kerapatan Probabilitas ini sebagai $p(x_1)$, tetapi ini adalah penyederhanaan dari notasi p yang lebih lama tetapi lebih jelas ($x = x_1$), yang menekankan bahwa adalah nilai skalar yang diberikan dengan mengevaluasi fungsi p pada $x = x_1$.

Aturan produk, kadang-kadang disebut sebagai "aturan dasar Probabilitas," menyatakan bahwa Probabilitas gabungan variabel acak A dan B dapat ditulis

$$P(A, B) = (P(A|B))P(B) \quad (2.1)$$

Aturan produk juga berlaku ketika A dan B adalah kelompok atau himpunan bagian dari peristiwa atau variabel acak. Aturan produk, kadang-kadang disebut sebagai "fundamental rule of probability" menyatakan bahwa Probabilitas gabungan variabel acak A dan B dapat ditulis.

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (2.2)$$



Gambar 2.1 Contoh *graph Bayesian Network*

Validitas langkah multiplikasi bergantung pada asumsi tunggal yaitu bahwa, nilai yang diberikan untuk setiap node parents, mengetahui nilai untuk setiap keturunan lain tidak mengubah Probabilitas yang terkait dengan setiap nilai yang mungkin. Dengan kata lain, perangkat non keturunan lainnya tidak memberikan informasi apa pun tentang kesamaan nilai node melebihi dan di atas informasi yang diberikan oleh parents. Ini dapat ditulis:

$$P(\text{node} | \text{parents plus any other descendant}) = P(\text{node} | \text{parents}) \quad (2.3)$$

yang harus berlaku untuk semua nilai node dan atribut yang terlibat. Dalam statistik sifat ini disebut *conditional independence* (independensi bersyarat). Perkalian adalah sah asalkan masing-masing simpul kondisional independen dari *grandparent*, *great grandparents*, dan set non keturunan lainnya, akan diberikan induknya [10]. Menerapkan aturan produk secara rekursif antara variabel tunggal dan variabel lainnya menimbulkan aturan lain yang dikenal sebagai *chain rule* yang menyatakan bahwa Probabilitas gabungan dari n atribut A_i dapat didekomposisi menjadi produk berikut:

$$P(A_1, A_2, \dots, A_n) = P(A_1) \prod_{i=1}^{n-1} P(A_{i+1} | A_i, A_{i-1}, \dots, A_1) \quad (2.4)$$

Penggandaan Probabilitas di Bayesian Networks mengikuti sebagai akibat langsung dari aturan berantai.

Dekomposisi berlaku untuk urutan atribut apa pun. Karena penggunaan BN adalah grafik asiklik, node dapat dipesan untuk memberikan semua leluhur dari simpul a_i yang indeksnya lebih kecil dari i . Kemudian, karena asumsi independensi bersyarat semua jaringan bayesian dapat ditulis dalam bentuk

$$P(A_1, A_2, \dots, A_n) = \prod_{i=1}^n P(A_i | \text{Parents}(A_i)) \quad (2.5)$$

Untuk variabel tanpa *parents*, kita perlu memperkirakan Probabilitas tanpa syarat :

$$P(A = a) = \frac{1}{N} \sum_{i=1}^N 1(A_i = a) \quad (2.6)$$

Di mana $1(A_i = a)$ hanyalah sebuah fungsi indikator mengembalikan i th ketika nilai yang diamati h untuk $A_i = a$ dan 0 sebaliknya. Perkiraan entri tabel Probabilitas bersyarat untuk $P(B | A)$ dapat dinyatakan menggunakan notasi yang mirip dengan prosedur penghitungan intuitif yang diuraikan dibawah[10] :

$$P(B = b | A = a) = \frac{P(B = b, A = a)}{P(A = a)} = \frac{\sum_{i=1}^N 1(A_i = a, B_i = b)}{\sum_{i=1}^N 1(A_i = a)} \quad (2.7)$$

Untuk kesederhanaan representasi kita menghilangkan subscript B selanjutnya. Jika X_i tidak memiliki parent, lokalnya distribusi Probabilitas dikatakan tidak bersyarat, jika tidak, itu bersyarat. Jika variabel diwakili oleh sebuah node diamati, maka node dikatakan sebagai node bukti, jika tidak, node dikatakan tersembunyi atau laten.

2.4 Kuliner

Istilah kuliner bersumber dari bahasa latin, yakni *culinarius*. *Culinarius* memiliki arti materi yang berkaitan dengan proses masak. Kata kuliner berasal dari komoditas yang berkaitan dengan masak-memasak dan kegiatan mengonsumsi

pangan. Kuliner bisa dimaknai sebagai pola pemakaian yang berdasarkan pada pangan atau hidangan. Wisata kuliner membuat makanan seperti subjek dan media. Makanan juga sebagai tujuan wisata dan instrumen bagi peningkatan pariwisata 156 Wisata kuliner diperlukan kerja sama semua panca indra seperti lidah, hidung, indra peraba, serta mata[11].

2.5 Youtube

Situs video YouTube sebagai salah satu bagian dari social networking dalam kategori media sosial dalam perkembangannya telah menghasilkan berbagai dampak nilai-nilai bagi para penggunanya. Nilai-nilai yang dimiliki oleh para penggunanya. Selain kemudahan dalam berhubungan satu dengan lainnya tanpa dipengaruhi jarak dan waktu. Progress dan kecepatannya dalam mencapai sebuah popularitas telah terwujud bagi para penggunanya[12].

Situs video YouTube dengan fungsinya yang dapat menciptakan hubungan interaksi dan komunikasi yang sangat dekat antar para penggunannya. Pada akhirnya akan tercipta sebuah situasi adaptasi antar budaya (Intercultural Adaptation) pada masyarakat atau komunitas pengguna yang pada akhirnya dapat menciptakan sebuah transisi satu budaya menjadi budaya yang baru. Secara teoritis diungkapkan oleh Sawyer (dalam Chen & Strarosta, 2005) bahwa dalam terdapat 4 (empat) tahapan melalui proses adaptasi antar budaya, yaitu: (a) Tahapan masa bulan madu (honeymoon), para masyarakat atau komunitas pengguna merasa sangat gembira (euphoria) dengan situasi yang dirasakan saat itu terhadap budaya barunya; (b) Tahapan masa krisis (crisis), merupakan sebuah situasi tidak nyaman dan frustrasi pada masyarakat atau komunitas pengguna yang di temukan dan dirasakan pada sisi lain dari budaya baru tersebut. Nilai-nilia budaya yang dirasakan tidak familiar; (c) Tahapan Penyesuaian (adjustment), merupakan sebuah situasi para masyarakat atau komunitas pengguna mulai mencoba untuk menyesuaikan diri dengan budaya barunya; (d) Tahapan bikulturalisme (biculturalism), merupakan tahapan akhir dimana para masyarakat atau komunitas pengguna mulai menyesuaikan diri dengan budaya barunya.

Situs video YouTube sebagai kategori dari media sosial secara mayoritas digunakan oleh para pengguna yang memiliki sifat extrovert dalam kehidupan sosialnya. Situs video YouTube oleh para pengguna extrovert digunakan sebagai media komunikasi meningkatkan frekuensi eksistensi dan hiburan bagi para pengguna lainnya dalam masyarakat media sosial. Seorang extrovert yang memiliki sifat keterbukaan yang besar umumnya ingin selalu menampilkan dominasinya dalam kehidupan sosial. Salah satunya melalui media sosial[12].

2.6 Web Scraping

Web Scraping adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya berupa halaman-halaman web dalam bahasa markup seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain.

Web scraping sering dikenal sebagai screen scraping. Web Scraping tidak dapat dimasukkan dalam bidang data mining karena data mining menyiratkan upaya untuk memahami pola semantik atau tren dari sejumlah besar data yang telah diperoleh. Aplikasi web scraping (juga disebut intelligent, automated, or autonomous agents) hanya fokus pada cara memperoleh data melalui pengambilan dan ekstraksi data dengan ukuran data yang bervariasi.

Teknik scraping dapat dilakukan dengan berbagai cara diantaranya menggunakan analisa html DOM (document object model) dan dengan menggunakan teknik pemrograman regular ekspresi. Kedua teknik ini mempunyai keunggulan tersendiri dan menghasilkan hasil yang tidak jauh berbeda. Pada teknik DOM dibutuhkan Xquery untuk mengekstrak konten utama dari halaman situs sedangkan pada teknik regular ekspresi ditentukan pola yang mengawali dan mengakhiri suatu konten utama pada halaman situs[13].

Pada penelitian yang dilakukan, web scraping menggunakan Youtube API(Application Program Interface) dengan menggunakan tools Youtube Comment Scraper.

2.7 Preprocessing

Preprocessing adalah tahapan untuk mempersiapkan teks menjadi data yang akan diolah di tahapan berikutnya. Teks yang akan dilakukan pada proses pada umumnya memiliki beberapa karakteristik, berdimensi tinggi, terdapat *noise* dan terdapat struktur yang tidak baik.

Adapun tahapan *preprocessing* yang akan dilakukan pada penelitian ini yaitu *case folding*, *convert emoticon*, *filtering*, *convert slangword*, *convert negation*, *stopword removal*, *stemming* dan *postag*.

2.7.1 Case Folding

Case folding merupakan proses yang dilakukan untuk menyeragamkan karakter pada data (dokumen/teks). Tahapan proses mengubah semua huruf dalam teks dokumen menjadi huruf kecil.

2.7.2 Convert Emoticon

Convert emoticon adalah proses mengkonversikan emoticon kedalam string yang sesuai dengan ekspresi emoticon itu sendiri. Convert emoticon dilakukan karena pada data komentar yang diambil dari *tag* 'kuliner' terdapat emoticon yang merupakan salah satu cara mengekspresikan persetujuan atau pertidaksetujuan dalam suatu komentar. Hal ini dirasa mempunyai pengaruh terhadap pengklasifikasian sentimen, oleh karena itu convert emoticon. Pada setelah menggunakan scraper untuk mengambil data yang ada di youtube, emoticon yang diambil berubah menjadi simbol-simbol. Rangkaian dari simbol itu bisa di konversikan menjadi sebuah kata yang memiliki makna untuk membantu menganalisis sentimen. Pada penelitian kali ini menggunakan *library* simbol dari Jakaton Femicidios yang terdapat 686 rangkaian simbol[6].

2.7.3 Filtering

Filtering digunakan untuk menyaring data. Data yang akan disaring berupa simbol-simbol yang sudah tidak dibuthkan lagi dalam proses analisis sentimen. Data yang akan diambil berupa alfabet yang berjumlah 26 karakter mulai dari huruf a-z.

2.7.4 *Convert Slangword*

Slang ragam bahasa tidak resmi dan tidak baku yang sifatnya musiman, dipakai oleh kaum remaja atau kelompok sosial tertentu untuk komunikasi intern dengan maksud agar yang bukan anggota kelompok tidak mengerti. Pada sosial media seperti youtube, data komentar terdapat banyak kata-kata yang tidak baku. Kata yang tidak baku akan mempengaruhi proses selanjutnya yang nantinya sulit dideteksi saat pelabelan di setiap katanya. Kata yang tidak baku tersebut akan diubah menjadi kata yang baku. Pada penelitian kali ini menggunakan fitur lokal, dimana kata-kata yang tidak baku yang terdapat dalam data akan diubah menjadi kata baku. Perubahan kata tidak baku menjadi kata baku didapat dari kitab gaul. Proses selanjutnya yaitu kata akan ditentukan kategorinya dan dilihat setiap *stemming* katanya yang menandakan bahwa kata itu sudah baku. Selanjutnya kata tersebut akan dianalisis melalui proses morfologi. Proses yang digunakan adalah milik *Andrew Castairs* dan *McCarthy*[14] yang gunanya juga untuk mendapatkan kata baku yang baru karena tidak semua kata tidak baku terdapat dalam kitab gaul untuk datanya terlampir di lampiran A.

Contoh : Kata tidak baku “coy” yang jika dilihat dari kitab gaul artinya adalah “kawan” termasuk dalam kategori nomina dan merupakan kata baku yang diuji dengan *stemming* yang artinya masih tetap “kawan” dan proses morfologinya yaitu membuat variasi baru.

2.7.5 *Convert Negation*

Convert Negation merupakan proses konversi kata-kata negasi yang terdapat pada suatu kalimat, karena kata negasi mempunyai pengaruh dalam merubah nilai sentimen pada sebuah kalimat. Di dalam klasifikasi teks sentimen, kata – kata negasi menjadi sangat berarti sehingga tidak boleh dihilangkan. Perbedaan arti “tak cinta” dan “tidak mudah” dapat menempatkan dokumen dalam kelas yang berbeda. Kata – kata negasi tersebut meliputi kata “bukan”, “tidak”, “tak” dan “tanpa”. Jika terdapat kata negasi maka akan digabungkan dengan kata setelahnya

2.7.6 *Stopword Removal*

Stopword Removal adalah proses untuk menghilangkan kata yang tidak relevan pada hasil *parsing* sebuah dokumen teks dengan cara membandingkan dengan *stoplist* yang ada. *Stoplist* berisi sekumpulan kata yang tidak relevan namun sering muncul dalam sebuah dokumen. *Stoplist* berisi sekumpulan *stopwords*.

Setiap kata akan diperiksa apakah masuk kedalam *stoplist* atau tidak, jika sebuah kata termasuk kedalam *stoplist* maka kata tersebut tidak akan diproses lebih lanjut dan akan dihilangkan. Sebaliknya jika sebuah kata tidak termasuk kedalam *stoplist* maka kata tersebut akan masuk ke proses berikutnya. Pada penelitian kali ini *stoplist* yang digunakan merupakan *stopword* dari *Library Sastrawi*.

2.7.7 *Stemming*

Steming merupakan salah satu tahapan dalam proses text pre-processing. Steming bertujuan untuk mentransformasikan kata menjadi kata dasarnya (root) dengan menghilangkan semua imbuhan kata (affixes) meliputi awalan kata (prefixes), sisipin kata (infixes), akhiran kata (suffixes) dan atau menghilangkan awalan dan akhiran kata (confixes) pada kata turunan. Salah satu library yang bisa digunakan dalam melakukan proses stemming bahasa indonesia adalah dengan menggunakan *Library Python Sastrawi* dimana *library* tersebut menerapkan algoritma *Algoritma Nazief* dan *Adriani*.

2.7.8 *Postag*

POS(*Part of speech*)tag merupakan suatu cara pengkategorian kelas kata, seperti kata benda, kata kerja, kata sifat dan sebagainya. POS Tagger merupakan sebuah aplikasi yang mampu melakukan proses anotasi part of speech tag untuk setiap kata didalam dokumen secara otomatis. Tools PoS Tag yang akan digunakan adalah library *postag CRF Tiger*.

2.7.9 *Bag of Word*

Semua dokumen dapat direpresentasikan secara sederhana menggunakan Bag-of-words (BoW). BoW adalah sebuah model yang merepresentasikan objek secara global misalnya kalimat teks atau dokumen sebagai bag (multiset) kata tanpa

memperdulikan tata bahasa bahkan urutan kata untuk menjaga keanekaragamannya. Dengan kata lain, BoW merupakan kumpulan kata-kata unik dalam dokumen. Contoh sederhana pembentukan bag-of-words untuk teks dokumen sebagai berikut

Teks: Sari senang membaca novel, Ina juga penggemar novel remaja.

Teks diatas dapat disusun menjadi BoW, dengan menggunakan kata unik yang direpresentasikan sekali saja sehingga membentuk urutan yang berbeda kemudian dihitung frekuensi kemunculannya.

Tabel 2.1 Contoh *Bag of Word*

| No | Kata | Distribusi Frekuensi |
|----|-----------|----------------------|
| 1 | Sari | 1 |
| 2 | Senang | 1 |
| 3 | Membaca | 1 |
| 4 | Novel | 2 |
| 5 | Ina | 1 |
| 6 | Juga | 1 |
| 7 | Penggemar | 1 |
| 8 | Remaja | 1 |

Distribusi frekuensi kata dapat dibandingkan dan digunakan untuk menilai kemiripan antara dua atau lebih dokumen dengan cara menghitung jarak keduanya [15].

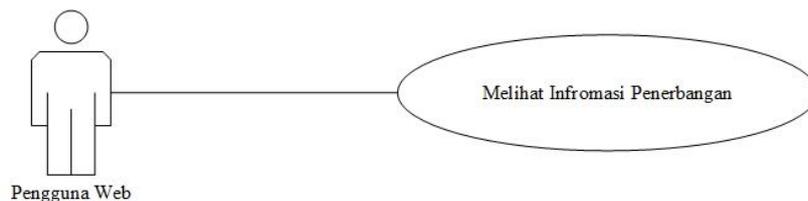
2.8 Unified Modeling Language

UML singkatan dari Unified Modeling Language merupakan bahasa pemodelan standar untuk merancang dan mendokumentasikan perangkat lunak dengan cara berorientasi objek. Ada beberapa diagram yang digunakan dalam proses pembuatan perangkat lunak berorientasi objek, diantaranya *use case diagram*, *activity diagram*, *class diagram* dan *sequence diagram*.

2.8.1 Use Case Diagram

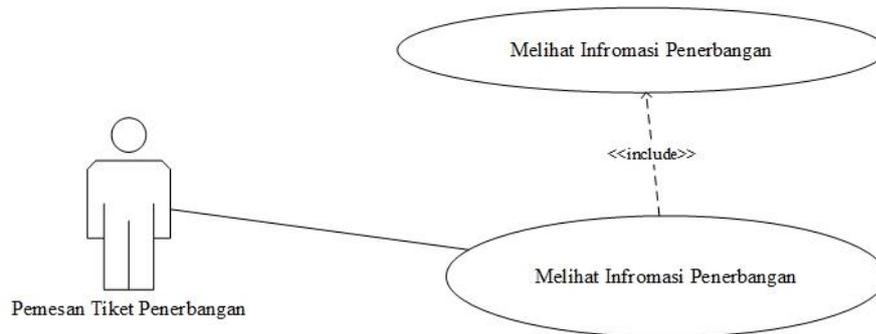
Use case diagram pada dasarnya digunakan untuk mendeskripsikan bagaimana entitas eksternal akan menggunakan sistem atau perangkat lunak. Entitas eksternal itu dapat berupa manusia atau sistem yang lain. Dalam diagram use case, entitas eksternal ini sering dinamakan sebagai actor. Deskripsi diagram use case ini lebih menekankan pada sistem dari sudut pandang penggunaanya dan juga menekankan pada interaksi yang terjadi di antara pengguna dengan sistem. Use case sangat membantu pengembang sistem untuk lebih jauh mendefinisikan ruang lingkup sistem serta batasan-batasannya.

Actor pada dasarnya adalah segala sesuatu yang berada di luar sistem atau perangkat lunak yang sedang dikembangkan. Setiap interaksi yang terjadi di antara actor dan sistem dimodelkan sebagai use case. Contoh diagram use case sederhana dapat dilihat pada Gambar 2.2.



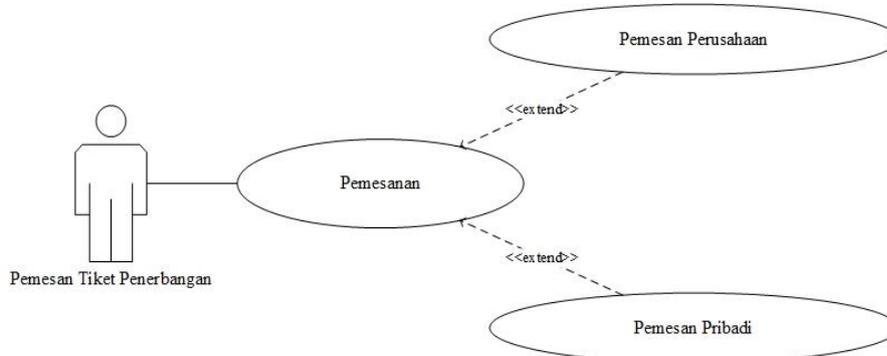
Gambar 2.2 Contoh use case

Dalam diagram use case, sering ditemukan asosiasi yang bertipe <include>. Asosiasi <include> ini menunjukkan bahwa use case tertentu, misalnya “Melihat Informasi Penerbangan” harus dilakukan terlebih dahulu sebelum use case lainnya seperti “Memesan Kursi Pesawat” dilakukan. Terdapat asosiasi yang terjadi di antara use case “Melihat Informasi Penerbangan” dan use case “Memesan Kursi Pesawat”, sedangkan pemesanan tiket penerbangan tidak dapat menggunakan use case “Memesan Kursi Pesawat” secara mandiri dari use case “Melihat Informasi Penerbangan”. Implementasi asosiasi <include> dalam kasus tersebut pada diagram use case dapat dilihat Gambar 2.3



Gambar 2.3 Contoh include use case

Selain asosiasi <include> dalam penggambaran juga dikenal asosiasi yang bertipe <extend>. Asosiasi <extend> ini hampir mirip dengan asosiasi bertipe <include>, hanya saja use case yang diasosiasikan sebagai <extend> tidak wajib dilaksanakan. Misalnya actor “Pemesan Tiket Penerbangan” dapat melakukan use case “Pemesanan” dengan terlebih dahulu melaksanakan use case “Pemesanan Perusahaan” atau “Pemesanan Pribadi”, tetapi tidak keduanya. Artinya, salah satu dari kedua use case ini harus dilaksanakan. Implementasi asosiasi <extend> dalam kasus tersebut pada diagram use case dapat dilihat Gambar



Gambar 2.4 Contoh extend use case

Salah satu model perancangan berorientasi objek yang saat ini paling sering digunakan di seluruh dunia adalah *Unified Modelling Language* (UML). UML sendiri adalah bahasa pemodelan untuk sistem atau perangkat lunak yang memiliki paradigma berorientasi objek.

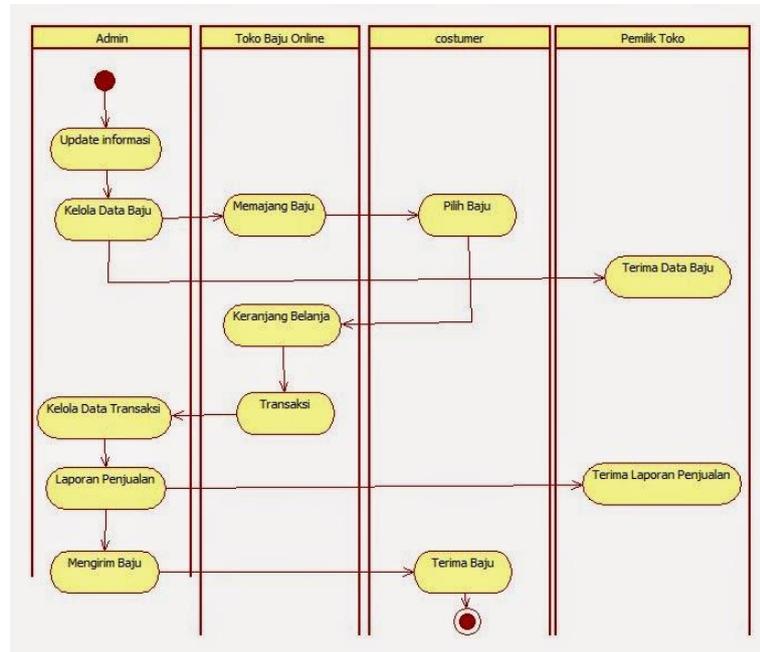
Unified Modelling Language (UML) adalah sebuah bahasa yang telah menjadi standar dalam industri untuk visualisasi, merancang dan mendokumentasikan sistem piranti lunak. UML menawarkan sebuah standar untuk

merancang model sebuah sistem. Dengan menggunakan UML, perancang atau pemodel dapat membuat model untuk semua jenis aplikasi piranti lunak, dimana aplikasi tersebut dapat berjalan pada piranti keras, sistem operasi dan jaringan apapun, serta ditulis dalam bahasa pemrograman apapun. UML lebih cocok diterapkan pada piranti berorientasi objek seperti C++, Java, C#, dan sebagainya. Tetapi UML juga tetap dapat digunakan untuk modeling aplikasi prosedural semisal VB atau C.

UML versi 2.0 mencakup 13 macam diagram dan perangkat yang berfungsi untuk menggambarkan sistem informasi berorientasi objek dengan sangat lengkap dan rinci. Meski demikian, tidak selalu ke-13 diagram dan perangkat tersebut digunakan saat para pengembang berupaya mengemabangkan perangkat lunak berorientasi objek. Beberapa diantara digaram UML yang digunakan adalah *Use Case Diagram*, *Activity Diagram*, *Sequence Diagram*, *Class Diagram*, dan *Collaboration Diagram*[15].

2.8.2 Activity Diagram

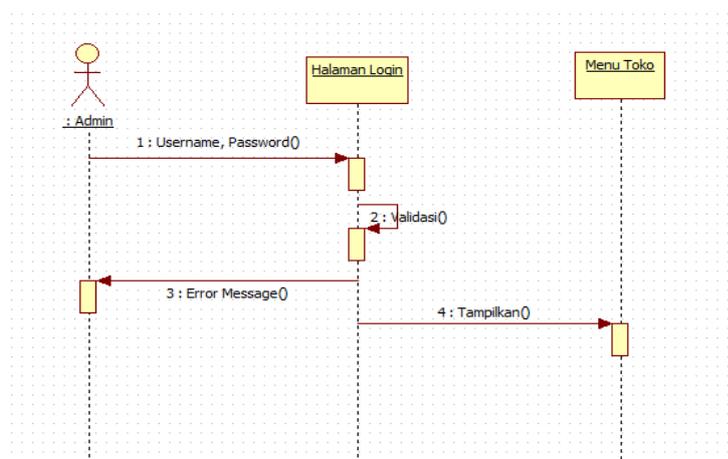
Sebuah *activity diagram* atau diagram aktivitas dapat direalisasikan oleh satu *use case* atau lebih. Aktivitas menggambarkan proses yang berjalan, sementara *use case* menggambarkan bagaimana actor menggunakan sistem untuk melakukan aktivitas. Sama seperti *state*, standar UML menggunakan segiempat dengan sudut membulat untuk menggambarkan aktivitas. *Decision* digunakan untuk menggambarkan behaviour pada kondisi tertentu. Untuk mengilustrasikan proses-proses paralel (*fork* dan *join*) digunakan titik sinkronisasi yang dapat berupa titik, garis horizontal atau vertikal. *Activity diagram* dapat dibagi menjadi beberapa *object swimlane* untuk menggambarkan objek mana yang bertanggung jawab untuk aktivitas tertentu. Contoh dari *activity diagram* dapat dilihat pada Gambar



Gambar 2.5 Contoh activity diagram

2.8.3 Sequence Diagram

Sequence diagram menjelaskan interaksi objek yang disusun berdasarkan urutan waktu. Secara mudahnya sequence diagram adalah gambaran tahap demi tahap, termasuk kronologi (urutan) perubahan secara logis yang seharusnya dilakukan untuk menghasilkan sesuatu sesuai dengan use case diagram. Contoh sequence diagram dapat dilihat pada Gambar.



Gambar 2.6 Contoh diagram sequence

2.9 Python

Python adalah bahasa pemrograman komputer, sama layaknya seperti bahasa pemrograman lain, misalnya C, C++, Pascal, Java, PHP, Perl dan lain-lain. Sebagai bahasa pemrograman, Python tentu memiliki varian, kosakata atau kata kunci, dan aturan tersendiri yang jelas berbeda dengan bahasa pemrograman lainnya.

Bahasa pemrograman Python disusun di akhir tahun 1980-an dan implementasinya baru dimulai pada Desember 1989 oleh Guido Van Rossum di Centrum Wiskunde & Information (CWI), sebuah pusat riset di bidang matematika dan sains, Amsterdam – Belanda; sebagai suksesor atau pengganti dari bahasa pemrograman pendahulunya, bahasa pemrograman ABC, yang juga dikembangkan di CWI oleh Leo Geurts, Lambert Meertens, dan Steven Pemberton[16].

Secara umum, para programmer banyak yang menjatuhkan pilihannya ke bahasa Python karena alasan – alasan berikut.

- a. Python memiliki konsep design yang bagus dan sederhana, yang berfokus pada kemudahan dalam penggunaan. Kode Python dirancang untuk mudah dibaca, dipelajari, digunakan ulang, dan dirawat. Selain itu, Python juga mendukung pemrograman berorientasi objek dan pemrograman fungsional.
- b. Python dapat meningkatkan produktivitas dan menghemat waktu bagi para programmer. Untuk memperoleh hasil program yang sama, kode Python jauh lebih sedikit dibandingkan dengan kode yang ditulis menggunakan bahasa–bahasa pemrograman lain.
- c. Program yang ditulis menggunakan Python dapat dijalankan di hampir semua sistem operasi (Linux, Windows, Mac, dan lain-lain), termasuk untuk perangkat-perangkat mobile.
- d. Python bersifat gratis atau bebas (*free*) dan *open-source*, meskipun digunakan untuk kepentingan komersial.

2.10 Confusion Matrix

Confusion matrix merupakan sebuah metode perhitungan yang digunakan untuk mencari keakuratan pada hasil klasifikasi[3]. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan klasifikasi yang seharusnya.

Pada pengukuran kinerja menggunakan *Confusion Matrix*, terdapat empat istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive* (TP), *True Negatif* (TN), *False Positive* (FP), dan *False Negatif* (FN). Nilai TP merupakan data positif yang terdeteksi dengan benar, nilai TN merupakan jumlah data negatif yang terdeteksi benar, nilai FP merupakan data negatif namun terdeteksi sebagai data positif, dan FN merupakan kebalikan dari nilai TP namun terdeteksi sebagai data negatif. Berikut contoh *confusion matrix* dapat dilihat pada Tabel 2.6 berikut.

Tabel 2.2 Aturan nilai dalam *confusion matrix*

| | | <i>Actual Class</i> | |
|------------------------|---|--------------------------|---------------------------|
| | | + | - |
| <i>Predicted Class</i> | + | <i>True Positive(TP)</i> | <i>False Positive(FP)</i> |
| | - | <i>False Negatif(FN)</i> | <i>True Negatif(TN)</i> |

Formula yang digunakan untuk klasifikasi multikelas adalah sebagai berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (2.8)$$

$$Precision = \frac{TP}{FP + TP} * 100\% \quad (2.9)$$

$$Recall = \frac{TP}{TP + FN} * 100\% \quad (2.10)$$