

## BAB II

### LANDASAN TEORI

#### 2.1 Data Mining

Menurut Jared Dean, Data Mining mengalami perkembangan yang pesat dengan kapasitas, variasi, nilai yg kecepataannya signifikan dari tahun ketahun. kapasitas dari data di buat melampaui batas jumlah yang akhir-akhir kebanyakan suatu organisasi kebingungan dalam membuat data dan nilai data tersebut. Di waktu yang sama data mining adalah penyimpanan data, *hardware* dengan kemampuan dan yang turun berubah drastis.

Sebagai data mining tidak step yang dimiliki secara garis besar memberikan koleksi beberapa step, teori, algoritma dan *hardware*. Semua komponen berubah dan tidak selalu sinkronasi dengan kenaikan kebutuhan data mining, machine learning dan *big data analytical* problem. Empat komponennya *disk, memory, central processing* dan *network* bisa menjadi 4 komponen pilar sebagai platform. Bagi yang memiliki stoll yang semua pilarnya harus memiliki kaki sama Panjang, dan lengan yang membuat frustrasi, berdiri dan pergi untuk mencari inspirasi. Hanya sistem data mining yang menjadi balace dan untuk penelitian dan components untuk memberikan pengguna pengalaman terbaik untuk meneliti sebuah problem.

Data mining pada skala apa pun tidak dapat dilakukan tanpa keahlian khusus perangkat lunak. Di dalam urutannya dalam menjelaskan evolusi dan progress dari data *hardware*, ada kebutuhan yang di buat untuk dalam skala yang lebih kecil dengan background yang menggunkan cara tradisional berinteraksi di antara hardware dan software. Data mining merupakan software paket dan berdiskusi di dalam detail di part pertama. Dahulu, cara tradisional data mining *software*

implementasi data mining loading data ke dalam memori dan satu rangkaian eksekusi dan kelebihan data. Oleh sebab jumlah *memory* nya di tab ketersedianya dan kecepatan dari processor nya. Ketika process bisa tidak sesuai dengan memory dan processnya bisa gagal. Satu rangkaian bisa juga gagal untuk mengambil kesempatan dari multicore server yang tidak digunakan sebagai multiple user. berada di sistem pada saat yang sama. Alasan utama kita melihat perubahan dramatis dalam data mining adalah related untuk di ubah di dalam storage teknologi seperti memudahkan komputersional kapabilitas. Namun, semua software paket tidak bisa memiliki ke untungan dari update kapasitas seperti hardware. Khusus nya kebenaran dari komputer model. Secara hari-hari evaluasi menjadi memastikan bahwa algoritma sudah terdistribusi dan sudah secara efektif memanfaatkan semua power computer yang tersedia. . (Jared Dean., 2014)

## **2.2 CRISP-DM (Cross-Industry Standard Process for Data Mining)**

Menurut Peter Chapman et al., CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah sebuah proses standar yang digunakan untuk mengatasi tantangan yang terjadi dalam proses data mining atau analisis data. Metodologi ini membantu dalam memberikan arahan langkah-langkah yang sistematis dalam mengumpulkan, memproses, dan menganalisis data dengan tujuan untuk menghasilkan wawasan yang berharga serta informasi yang relevan.

CRISP-DM terdiri dari enam tahap utama, yaitu:

1. Business Understanding: Fokus utamanya adalah pada pemahaman terhadap tujuan bisnis dan masalah yang ingin dipecahkan. Kerjasama dengan para pemimpin bisnis sangatlah penting untuk memahami kebutuhan bisnis serta menetapkan tujuan analisis data yang ingin dicapai.

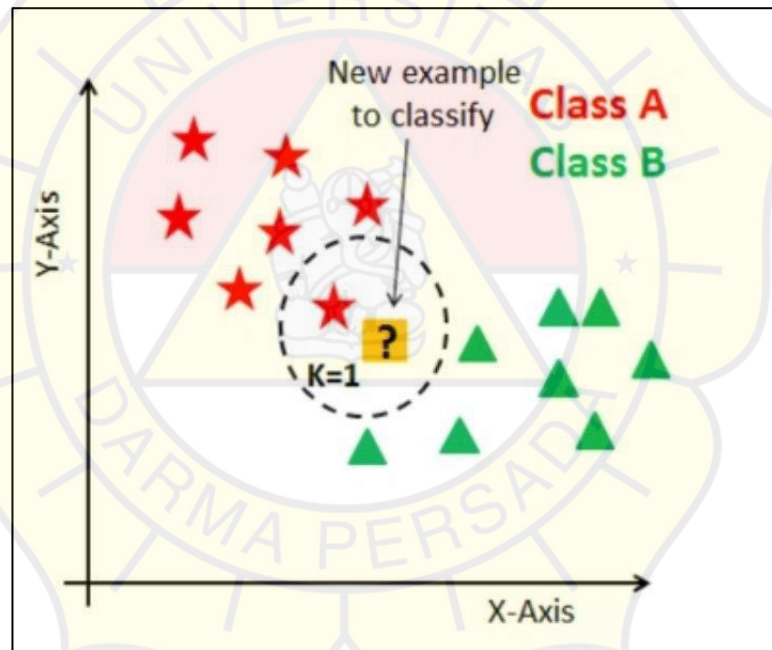
2. Data Understanding: Eksplorasi data untuk memahami sumber data yang tersedia, jenis data yang ada, dan potensi masalah atau kekurangan dalam data. Pada tahap ini, tim melakukan pemeriksaan kualitas data.
3. Data Preparation : Proses transformasi data dan seleksi variabel dilakukan untuk mengatasi kekurangan data dan meningkatkan kualitas data.
4. Modeling: Tahap ini melibatkan pengembangan dan evaluasi model analisa data yang sesuai dengan tujuan bisnis. Metode analisis data yang digunakan dapat bervariasi, seperti regresi, klasifikasi, atau klustering.
5. Evaluation: Pada tahap ini, model yang telah dikembangkan dievaluasi dan divalidasi untuk memastikan bahwa model tersebut dapat menghasilkan hasil yang akurat dan relevan. Jika diperlukan, perbaikan atau penyesuaian model dapat dilakukan.
6. Deployment: Tahap terakhir adalah penerapan hasil analisis data dalam lingkungan produksi. Hasil analisis dapat disampaikan kepada pemangku kepentingan untuk mendukung pengambilan keputusan bisnis. (Peter Chapman et al.,)

### **2.3 K-Nearest Neighbors**

Menurut Ferdinan Kusumah et al., Algoritma K-nearest Neighbors merupakan metode supervised learning di mana hasil dari instance yang baru ditentukan berdasarkan mayoritas dari kategori k-tetangga terdekat. Tujuan utama dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan sampel dari data latih. Algoritma k-Nearest Neighbor menggunakan klasifikasi tetangga terdekat sebagai nilai prediksi dari nilai instance yang baru.

Dalam KNN, nilai K menunjukkan jumlah tetangga terdekat yang akan digunakan dalam proses pengklasifikasian. Jumlah tetangga ini merupakan faktor kunci dalam algoritma ini. Biasanya, nilai K adalah bilangan ganjil, terutama jika jumlah kelasnya adalah 2.

Ketika nilai K sama dengan 1, maka algoritma tersebut dikenal sebagai algoritma tetangga terdekat. Ini merupakan kasus yang paling sederhana. Misalnya, jika P1 merupakan titik yang akan diprediksi labelnya, algoritma akan mencari satu titik terdekat dari P1, dan kemudian label dari titik terdekat tersebut akan ditugaskan kepada P1



**Gambar 2.4** Klasifikasi KNN (Sumber Ferdinan Kusumah et al., , 2021)

Misalkan P1 merupakan inti dari proses prediksi label. Pertama, langkahnya adalah menemukan titik terdekat dengan P1, dan setelah itu, melakukan klasifikasi poin dengan menggunakan mayoritas suara dari tetangganya. Setiap objek memberikan suara untuk kelasnya masing-masing, dan kelas dengan suara terbanyak akan diambil sebagai prediksi akhir. Untuk menemukan titik-titik serupa

yang terdekat, prosesnya melibatkan perhitungan jarak antara titik-titik tersebut menggunakan berbagai ukuran jarak seperti Euclidean, Hamming, dan Minkowski.

Rumus Euclidean Distance :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

- a.  $d(x,y)$  adalah jarak antara data x ke data y
- b.  $x_1$  adalah data testing ke-i
- c.  $y_1$  adalah data training ke-i

Dalam algoritma k-NN, langkah-langkahnya adalah sebagai berikut:

1. Tentukan nilai parameter k, yang merupakan jumlah tetangga terdekat yang akan digunakan dalam proses klasifikasi (nilai k biasanya dipilih secara manual).
2. Hitung jarak antara setiap data training dan data testing, dengan menggunakan metode Euclidean distance yang merupakan pilihan umum dalam banyak penelitian.
3. Urutkan data training berdasarkan jarak terkecil dari data testing.
4. Tetapkan kelas untuk data testing, di mana kelas yang dipilih adalah kelas dengan jumlah terbanyak dari nilai k tetangga terdekat pada data testing tersebut. (Ferdinan Kusumah et al., 2021)

## 2.4 Sistem *Fashion*

Menurut Dr. Sutisna & Prof. Dr. Marcellia Susan, banyak orang umumnya mengasosiasikan fashion dengan pakaian, baik itu pakaian pria maupun wanita. Namun, dalam konteks ini, konsep desain tidak terbatas hanya pada pakaian, tetapi juga mencakup berbagai produk yang memiliki makna dan simbol kebudayaan.

Sistem desain melibatkan individu dan organisasi yang terlibat dalam menciptakan makna simbolis dan mengemasnya dalam bentuk barang-barang konsumsi. Produk desain merupakan hasil dari sistem desain yang mencakup berbagai fenomena budaya, seperti musik, seni, arsitektur, dan ilmu pengetahuan. Bahkan, praktik bisnis pun dapat dikategorikan sebagai bagian dari proses desain.

Sistem fashion terus berkembang di masyarakat, yang dapat menghasilkan budaya baru dan bahkan budaya populer. Sistem fashion yang berkembang dengan cepat cenderung menghasilkan budaya populer lebih banyak. Budaya tinggi biasanya sulit digantikan oleh budaya populer, meskipun budaya populer juga memiliki tempatnya sendiri. Sebagai contoh, musik klasik, sebagai representasi dari budaya tinggi, tetap eksis hingga saat ini meskipun musik pop terus bermunculan.

Budaya populer cenderung mengalami proses lupakan, di mana pengikutnya akan melupakan budaya yang sudah menjadi kuno saat mereka beralih ke budaya populer yang baru. Namun, suatu saat nanti, budaya populer yang tenggelam tersebut dapat kembali dihidupkan oleh orang-orang yang tertarik padanya.

## **2.5 E-commerce Management**

Menurut Dave Chaffey, dalam konteks penamaan transaksi online, terdapat dua pendekatan yang umum digunakan, yaitu "bisnis digital" dan "e-commerce". Meskipun kedua istilah ini sering kali dianggap sama oleh sebagian orang, namun bagi yang lainnya, keduanya memiliki perbedaan yang signifikan. Yang paling penting adalah konsistensi dalam penggunaan istilah ini di dalam suatu organisasi, sehingga karyawan dan pemangku kepentingan eksternal memahami dengan jelas bagaimana organisasi tersebut memanfaatkan komunikasi elektronik.

Perbedaan yang ditekankan dalam buku ini adalah bahwa "perdagangan elektronik" (e-commerce) merujuk pada segala jenis tindakan elektronik antara organisasi dan pemangku kepentingan, termasuk transaksi keuangan, pertukaran informasi, atau layanan lainnya. Transaksi e-commerce dapat berupa transaksi jual beli (e-commerce sisi beli atau sisi jual), dan masalah manajemen yang terkait dengan masing-masing aspek tersebut dibahas secara terpisah di dalam bagian kedua buku ini.

Di sisi lain, istilah "bisnis digital" digunakan sebagai konsep yang lebih luas, yang mencakup e-commerce serta segala bentuk transaksi elektronik dalam suatu organisasi. Manajemen e-commerce melibatkan penentuan prioritas aktivitas e-commerce sisi beli dan sisi jual, serta merencanakan dan mengalokasikan sumber daya untuk mendapatkan manfaat yang diinginkan. Rencana ini harus difokuskan pada pengelolaan berbagai risiko kegagalan, yang dapat bervariasi mulai dari masalah teknis seperti transaksi yang gagal atau situs yang lambat, hingga masalah dengan layanan atau pemenuhan pelanggan.

Perkembangan media sosial dan interaksi peer-to-peer di situs web perusahaan, blog, komunitas, dan jejaring sosial telah mengubah dinamika perdagangan online. Demikian pula, adopsi teknologi seluler oleh konsumen melalui situs dan aplikasi seluler menawarkan platform baru untuk berinteraksi dengan pelanggan yang harus dievaluasi dan diprioritaskan. Menentukan teknologi dan pendekatan pemasaran yang akan diprioritaskan merupakan tantangan bagi semua organisasi dalam era digital ini.

## **2.6 Pemrograman Dasar**

### **2.6.1 HTML (*Hyperlink Markup Language*)**

Menurut Steven Suehering & Janet Valade, HyperText Markup Language (HTML) adalah bahasa yang digunakan dalam pengembangan web. Saat Anda mengakses halaman web melalui browser seperti Internet Explorer, Firefox, atau Safari, browser tersebut mengunduh dan menampilkan dokumen HTML. Secara sederhana, HTML adalah jenis dokumen, mirip dengan dokumen yang Anda buat di pengolah kata. Program seperti Microsoft Word digunakan untuk melihat dokumen pengolah kata agar dapat membaca dan menampilkannya.

Dalam konteks web, browser web berfungsi sebagai program yang tahu cara membaca dan menampilkan dokumen yang dibuat dengan HTML. Dokumen pengolah kata dapat dibuat dan dibaca dengan satu program, sedangkan dokumen HTML memerlukan program berbeda untuk pembuatannya dan pembacaannya. Anda tidak dapat membuat dokumen HTML menggunakan browser. Untuk membuat dokumen HTML, Anda perlu menggunakan program yang disebut editor. Editor tersebut dapat sederhana seperti Notepad yang disertakan dengan Microsoft Windows, atau lebih kompleks seperti Eclipse atau Microsoft Visual Studio. Biasanya, Anda dapat menggunakan program yang sama untuk membuat dokumen HTML dan program PHP. (Steven Suehering & Janet Valade, 2013 )

### **2.6.2 CSS (*Cascading Style Sheet*)**

Menurut Steven Suehering & Janet Valade, CSS (Cascading Style Sheets) adalah alat yang melengkapi HTML dengan memberikan tampilan dan nuansa pada halaman web. Secara default, halaman HTML cenderung polos, menggunakan font default dan ukuran font standar. Dengan menggunakan CSS, Anda dapat



meningkatkan penampilan halaman tersebut dengan menambahkan warna, gambar latar belakang, mengubah jenis dan ukuran font, serta menambahkan dekorasi seperti garis batas di sekitar area konten.

Meskipun CSS memiliki bahasa tersendiri yang terpisah dari HTML, namun penggunaan CSS tidak terpisahkan dari HTML. Dengan kata lain, meskipun HTML dapat berdiri sendiri dan menyajikan halaman ke browser, namun CSS tidak dapat berfungsi tanpa adanya halaman HTML. Anda tidak akan menulis halaman CSS secara mandiri; sebaliknya, Anda akan menulis kode HTML terlebih dahulu dan kemudian menggunakan CSS untuk memperindah dan menata halaman tersebut agar sesuai dengan keinginan Anda.

### **2.6.3 *Javascript***

Menurut Steven Suehering & Janet Valade, JavaScript telah menjadi bagian integral dari halaman web modern. Saat Anda menggunakan layanan seperti Google Maps dan dapat menggulir peta hanya dengan menyeret, atau saat Anda mencari detail penerbangan dan situs secara otomatis menyarankan bandara saat Anda mulai mengetik, itu semua berkat JavaScript. Berbagai widget dan peningkatan fungsionalitas yang kita temui saat menjelajahi web sebenarnya merupakan hasil dari program-program JavaScript yang bekerja di latar belakang.

Program JavaScript dijalankan di browser web pengguna, yang menghadirkan tantangan tersendiri. Di satu sisi, menjalankan JavaScript di browser membebaskan server dari beban pemrosesan program. Namun, di sisi lain, karena berjalan di berbagai versi browser, program JavaScript bisa memiliki perilaku yang sedikit berbeda tergantung pada versi browser yang digunakan pengguna. Bahkan, pengguna juga memiliki kemampuan untuk menonaktifkan JavaScript sepenuhnya,

yang dapat memengaruhi fungsionalitas dari situs web yang menggunakan JavaScript.

### **2.6.3 PHP (*Hypertext Processor*)**

Menurut Steven Suehering & Janet Valade, PHP merupakan bahasa scripting yang didesain khusus untuk digunakan dalam lingkungan web. PHP memiliki berbagai fitur yang membantu dalam memproses tugas-tugas pemrograman yang diperlukan untuk mengembangkan aplikasi web dinamis.

Perangkat lunak PHP beroperasi bersama dengan server web, yang bertanggung jawab untuk mengirimkan halaman web kepada pengguna di seluruh dunia. Saat pengguna mengetikkan URL ke dalam bilah alamat browser web, mereka sebenarnya mengirimkan pesan ke server web pada URL tersebut, meminta server untuk mengirimkan file HTML yang terkait. Kemudian, server web merespons permintaan tersebut dengan mengirimkan file yang diminta. Browser web kemudian membaca file HTML dan menampilkan halaman web yang sesuai.

Ketika pengguna mengklik tautan di halaman web, mereka kembali meminta file dari server web. Server web akan memproses permintaan ini, seperti juga saat pengguna mengklik tombol pada halaman web yang mengirimkan formulir. Proses ini berlangsung serupa ketika PHP digunakan. Ketika file diminta, server web secara otomatis menjalankan PHP, yang kemudian menghasilkan HTML untuk dikirimkan kembali ke browser, sesuai dengan logika pemrograman yang terkandung dalam skrip PHP.

### **2.6.4 MYSQL**

Menurut Steven Suehering & Janet Valade, setiap situs web dinamis membutuhkan basis data sebagai backend. Basis data ini dapat berisi informasi

yang ditampilkan oleh halaman web kepada pengguna, atau mungkin bertujuan untuk menyimpan informasi yang diberikan oleh pengguna. Dalam beberapa kasus, basis data berfungsi sebagai penyimpanan informasi yang tersedia serta menerima dan menyimpan informasi baru.

MySQL, yang merupakan basis data yang paling umum digunakan untuk situs web, telah dikembangkan agar cepat dan ringan, terutama untuk keperluan situs web. MySQL menjadi sangat populer untuk digunakan bersama dengan situs web yang menggunakan PHP, karena PHP dan MySQL dapat berintegrasi dengan baik.

Perangkat lunak MySQL terdiri dari beberapa komponen, termasuk server MySQL dan sejumlah program utilitas yang membantu dalam administrasi database. Inti dari sistem MySQL adalah server MySQL itu sendiri, yang berperan sebagai pengelola sistem database. Server MySQL menangani semua instruksi yang terkait dengan database. Misalnya, jika Anda ingin membuat basis data baru, Anda akan mengirimkan pesan ke server MySQL untuk memberitahukan instruksi tersebut. Server MySQL akan menanggapi dengan membuat subdirektori baru di direktori data, memberikannya nama yang sesuai, dan menyimpan file-file yang diperlukan dalam subdirektori tersebut dengan format yang tepat.

Sebelum menerima instruksi dari pengguna, server MySQL harus berjalan dan siap menerima permintaan. Biasanya, server MySQL diatur untuk mulai berjalan saat komputer dinyalakan dan beroperasi secara terus-menerus. Namun, ini dapat diatur ulang untuk dimulai secara manual sesuai kebutuhan. Saat beroperasi, server MySQL terus memantau pesan yang diterimanya dan menjalankan instruksi sesuai permintaan.

## 2.7 Behavior Diagram

### 2.7.1 *Diagram Use Case (Use Case Diagram)*

Menurut Martina Seild et al., suatu use case merupakan gambaran tentang fungsionalitas yang diharapkan dari sistem yang sedang dikembangkan. Use case ini mencakup serangkaian fungsi yang dijalankan saat menggunakan sistem tersebut. Suatu use case memberikan manfaat yang nyata bagi satu atau lebih aktor yang terlibat dalam interaksi dengan use case tersebut. Diagram use case tidak menggambarkan struktur internal atau implementasi sebenarnya dari suatu use case. Secara umum, suatu use case diaktifkan baik oleh pemanggilan dari seorang aktor atau oleh suatu trigger event, yang disebut sebagai pemicu. Contoh dari pemicu tersebut adalah berakhirnya semester, yang mengakibatkan use case untuk penerbitan sertifikat harus dijalankan.

Kasus penggunaan ditentukan dengan mengumpulkan keinginan pelanggan dan menganalisis masalah yang ditentukan dalam bahasa alami ketika ini menjadi dasar analisis persyaratan. Namun, kasus penggunaan juga dapat digunakan untuk mendokumentasikan fungsionalitas yang ditawarkan sistem. Sebuah use case biasanya direpresentasikan sebagai elips. Nama use case ditentukan langsung di dalam atau tepat di bawah elips. Sebagai opsi alternatif, sebuah use case bisa direpresentasikan dengan sebuah persegi panjang yang berisi nama use case di bagian tengahnya, dengan elips kecil diletakkan di sudut kanan atas. Semua alternatif sama validnya, namun alternatif pertama, elips yang berisi nama use case, adalah yang paling umum digunakan. (Martina Seild et al., 2012)

## 2.7.2 Activity Diagram

Menurut Martina Seild et al., diagram aktivitas bertujuan untuk memodelkan proses pemrosesan prosedural dalam suatu sistem. Diagram ini menggambarkan aliran kontrol dan aliran data antara langkah-langkah tindakan yang diperlukan untuk melaksanakan suatu aktivitas.

Dalam UML, diagram aktivitas menggunakan konsep bahasa berorientasi aliran yang terinspirasi dari bahasa untuk mendefinisikan proses bisnis. Mereka juga mengadopsi konsep-konsep yang ada untuk menggambarkan proses komunikasi secara bersamaan, seperti konsep token jaring Petri. Fitur khusus dari diagram aktivitas adalah kemampuannya untuk memodelkan sistem berorientasi objek dan non-berorientasi objek, sehingga memungkinkan penentuan aktivitas secara independen dari objek. Ini berarti Anda dapat memodelkan fungsi pustaka serta proses bisnis dan organisasi dunia nyata.

Standar UML tidak menetapkan bentuk notasi spesifik apa pun untuk aktivitas. Selain elemen notasi berbasis aliran pada diagram aktivitas, standar ini juga mengizinkan bentuk notasi lain, seperti diagram struktural atau bahkan kode semu. Sejumlah aliran kontrol berulang dan pola aliran data telah muncul selain elemen notasi khusus. Mereka digunakan khususnya untuk memodelkan proses bisnis dan telah terbukti sangat berguna untuk proses yang kompleks. Konstruksi ini disebut sebagai "pola alur kerja". Untuk gambaran umum tentang jenis pola ini serta panduan tentang cara memodelkan pola berdasarkan konsep diagram aktivitas .

Dalam bab ini, pembaca yang jeli akan memperhatikan bahwa tidak semua contoh memodelkan proses yang lengkap-beberapa model dibatasi pada ekstrak

proses. Jadi, misalnya, beberapa diagram tidak memuat node awal dan akhir. Namun dalam praktiknya, diagram aktivitas yang lengkap harus memiliki titik awal dan akhir yang jelas. (Martina Seild et al., 2012)

### **2.7.3 Sequence Diagram**

Menurut Martina Seild et al., Interaksi menentukan bagaimana pesan dan data dipertukarkan antara mitra interaksi. Mitra interaksinya bisa berupa manusia, seperti dosen atau mahasiswa, atau non-manusia, seperti server, printer, atau perangkat lunak yang dapat dieksekusi. Interaksi dapat berupa percakapan antara banyak orang-misalnya, ujian lisan. Alternatifnya, interaksi dapat memodelkan protokol komunikasi seperti HTTP atau mewakili pertukaran pesan antara manusia dan sistem perangkat lunak-misalnya, antara dosen dan sistem administrasi mahasiswa ketika dosen mempublikasikan hasil ujian. Interaksi juga dapat berupa rangkaian pemanggilan metode dalam suatu program atau sinyal seperti alarm kebakaran dan proses komunikasi yang dihasilkan.

Interaksi menggambarkan interaksi antara beberapa mitra interaksi dan terdiri dari serangkaian pesan. Pengiriman atau penerimaan suatu pesan dapat dipicu oleh terjadinya peristiwa tertentu, misalnya diterimanya pesan lain, dan dapat terjadi pada waktu-waktu tertentu, misalnya pukul 05.00. Batasan yang telah ditentukan sebelumnya menentukan prasyarat apa pun yang diperlukan yang harus dipenuhi agar interaksi berhasil. Misalnya melanjutkan proses komunikasi yang telah diuraikan di atas, dosen harus login ke sistem sebelum memasukkan nilai mahasiswanya.

Di UML, Anda menggunakan diagram interaksi untuk menentukan interaksi. Dalam diagram interaksi, Anda selalu memodelkan skenario konkret,

artinya pertukaran pesan terjadi dalam konteks tertentu untuk memenuhi tugas tertentu. Interaksi biasanya hanya menggambarkan bagian tertentu dari suatu situasi. Seringkali ada jalur eksekusi valid lainnya selain interaksi. (Martina Seild et al., 2012).





**TEKNOLOGI INFORMASI**  
**UNIVERSITAS DARMA PERSADA**