

BAB 2

TINJAUAN PUSTAKA

2.1. Sejarah Universitas Komputer Indonesia

Universitas Komputer Indonesia merupakan sebuah perguruan tinggi swasta bernama Lembaga Pendidikan Komputer Indonesia Jerman (LPKIG) pada tahun 1994 dan berkantor pusat di Jl. Dipatiukur 102 Bandung, Jawa Barat. Lembaga ini telah meluncurkan program pelatihan satu tahun yang terdiri dari lima program studi: Ahli Komputer Aplikasi Bisnis, Ahli Komputer Keuangan dan Perbankan, Ahli Komputer Akuntansi dan Perpajakan, Ahli Komputer Manajemen dan Pemasaran, dan Sekretaris Eksekutif. Pada tahun 1995 dibuka jenjang pelatihan tiga tahun dengan program studi baru: Ahli Komputer Teknik Informatika, Ahli Komputer Manajemen Informatika, dan Sekretaris Eksekutif.

Pada tanggal 24 Desember 1998 dibentuklah Yayasan Science dan Teknologi dan dilanjutkan dengan pengajuan pendirian STIMIK IGI dan STIE IGI ke DIKTI. Pada bulan Juli 1999 STIE IGI diresmikan dengan keluarnya SK Mendiknas no. 119/D/O/1999 dengan lima program studi yaitu Akuntansi S1, Manajemen S1, Manajemen Pemasaran D3, Keuangan Perbankan D3 serta Akuntansi D3. Pada bulan Agustus 1999 STIMIK IGI diresmikan dengan keluarnya SK Mendiknas no. 143/D/O/1999 dengan lima program studi yaitu Teknik Informatika S1, Manajemen Informatika D3, Teknik Komputer D3, Komputerasi Akuntansi D3 serta Teknik Informatika D3.

Agar Sistem Pendidikan lebih Efisien, Efektif, Produktif dengan Struktur Organisasi yang lebih baik, enam bulan kemudian dilakukan usulan ke DIKTI untuk melakukan Merger kedua Sekolah Tinggi diatas menjadi Universitas. Pada tanggal 8 Agustus 2000 keluarlah SK MENDIKNAS no. 126/D/O/2000 atas Universitas Komputer Indonesia yang disingkat dengan nama UNIKOM. Pada SK tersebut sekaligus diijinkan dibukanya 11 program studi baru yaitu Teknik Komputer S1, Manajemen Informatika S1, Teknik Industri S1, Teknik Arsitektur S1, Perencanaan Wilayah dan Kota S1, Ilmu Hukum S1, Ilmu

Komunikasi S1, Ilmu Pemerintahan S1, Desain Interior D3, Desain Komunikasi Visual S1 dan Desain Komunikasi Visual D3.

2.2. Logo Universitas



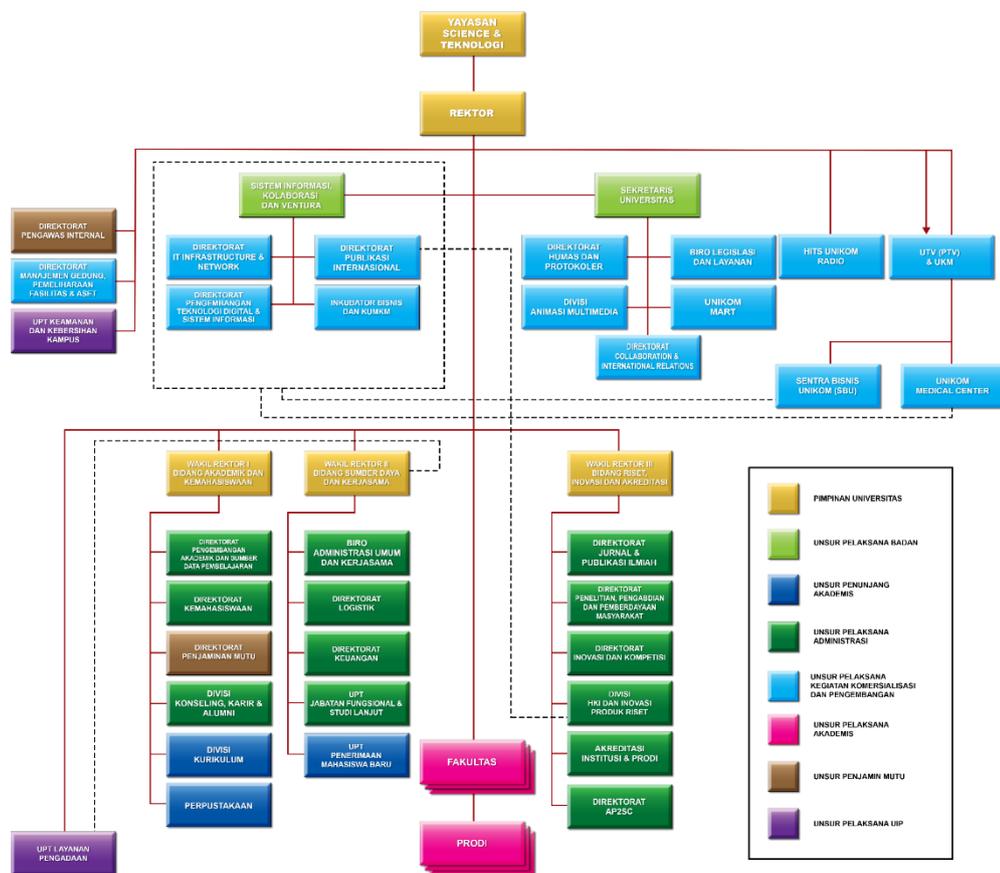
Gambar 2.1 Logo Universitas Komputer

Logo Universitas Komputer Indonesia terdiri dari 8 bagian yaitu :

- a. **Bingkai Segi Lima** : Melambangkan UNIKOM berlandaskan falsafah negara yakni Pancasila dan Undang-Undang Dasar 1945.
- b. **Lingkar Dalam Segi Lima Tempat Tulisan Berwarna Kuning** : Melambangkan motto UNIKOM menuju kejayaan yakni Quality Is Our Tradition.
- c. **Bulatan Dalam Berwarna Biru** : Melambangkan UNIKOM bertujuan menghasilkan ilmuwan unggul dan berpikiran maju yang Bertaqwa kepada Tuhan Yang Maha Esa.
- d. **Komputer** : Melambangkan ciri utama UNIKOM yang memberikan pendidikan Teknologi Informasi dan Komputasi pada seluruh Jurusan yang ada dilingkungan Universitas Komputer Indonesia, menjadi Universitas Terdepan dibidang Teknologi Informasi dan Komputer serta sebagai Universitas komputer pertama di Indonesia.
- e. **Stasiun Relay** : Melambangkan UNIKOM menyelenggarakan Pendidikan Tinggi kearah masyarakat industri maju dengan sistem pendidikan yang kondusif dan tenaga pengajar berkualitas untuk menghasilkan lulusan-lulusan terbaik.

- f. **Satelit** : Melambungkan UNIKOM berwawasan Global dan menjadi pusat unggulan dibidang IPTEK & seni yang mendukung Pembangunan Nasional serta berorientasi pada kepentingan masyarakat, bangsa dan negara.
- g. **Cakrawala** : Melambungkan indahnya menggapai Cita-cita dan mengejar ilmu setinggi Langit.
- h. **Buku** : Melambungkan sumber ilmu yang tiada habis-habisnya.

2.3. Struktur Organisasi



Gambar 2.2 Struktur Organisasi

2.4. Visi dan Misi

2.4.1. Visi

Menjadi Universitas Terkemuka dibidang Teknologi Informasi & Komunikasi, berwawasan global, berjiwa entrepreneur dan menjadi pusat di

bidang ilmu pengetahuan dan teknologi yang mendukung Pembangunan nasional serta berorientasi pada kepentingan Masyarakat.

2.4.2. Misi

Menyelenggarakan Pendidikan tinggi modern berdasarkan budaya organisasi UNIKOM, PIQIE (Professionalism, Integrity, Quality, Information Technology, Excellence), dengan system Pendidikan yang kondusif dan program – program studi yang berbasis pada software (Perangkat Lunak), Hardware (Perangkat Keras), Entrepreneurship (Kewirausahaan), dan animasi multimedia dengan mengoptimalkan sumber daya yang ada berdasarkan prinsip efisensi, efektifitas, dan produktifitas

2.5. Landasan Teori

Landasan teori ini berisikan teori-teori pendukung yang digunakan dalam proses analisis dan implementasi pada permasalahan yang diangkat dalam penelitian ini.

2.5.1. Promosi

Promosi dilakukan oleh suatu perusahaan untuk menunjukkan keberadaan produk dan memberi pembeli keyakinan bahwa produk tersebut akan bermanfaat. Salah satu metode untuk meningkatkan jumlah penjualan adalah promosi[3]. Semua alat yang termasuk dalam bauran promosi termasuk dalam promosi, yang fungsi utamanya adalah komunikasi yang membujuk[4].

2.5.2. Data

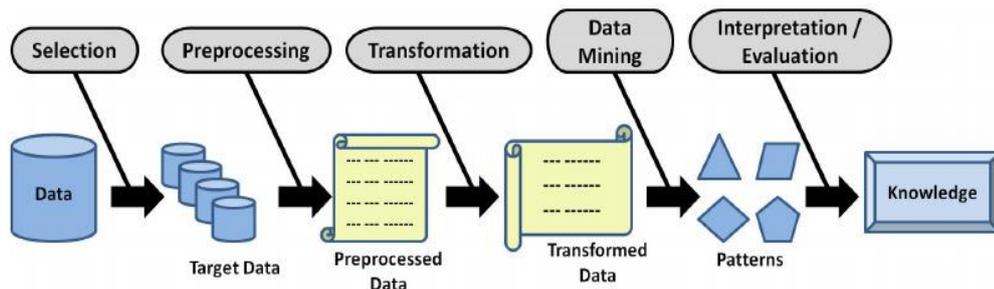
Data merupakan Kumpulan fakta atau sesuatu digunakan sebagai *input* yang diolah dalam proses dan akan menghasilkan suatu informasi (*output*). Sebuah data harus sesuai kebenarannya (*reliable*), akurat, tepat waktu dan mencakup ruang lingkup yang luas[5].

Jenis-jenis data dapat dibagi menurut sifatnya , menurut sumbernya, menurut cara memperolehnya dan menurut waktu pengumpulannya[6].

Meurut sifatnya data dibagi menjadi dua jenis, yaitu data kualitatif (non-metrik) dan data kuantitatif (numerik). Data kualitatif dapat disebut data yang bukan berupa angka. Data kuantitatif adalah data berupa angka dalam arti sebenarnya. Jenis data menurut sumbernya yang didasarkan pada sumber perolehan data tersebut dibagi menjadi dua, yaitu data internal dan data eksternal. Data internal adalah data yang dikumpulkan oleh suatu organisasi untuk menggambarkan keadaan atau kegiatan organisasi yang bersangkutan. Data eksternal adalah data yang dikumpulkn untuk menggambarkan keadaan atau kegiatan di luar organisasi tersebut. Data berdasarkan cara memperolehnya dibagi menjadi dua, yaitu data primer dan data sekunder. Data primer adalah data yang dikumpulkan dan diolah sendiri oleh organisasi atau perorangan langsung dari objeknya. Data sekunder adalah data yang diperoleh dalam bentuk jadi dan telah diolah oleh pihak lainnya. Data berdasarkan waktu pengumpulannya dibedakan menjadi dua jenis, yaitu data *cross section* dan data berkala (*time series*). Data *cross section* adalah data yang dikumpulkan dalam suatu periode tertentu. Data berkala (*time series*) adalah data yang dikumpulkan dari waktu ke waktu

2.5.3. Knowledge Discovery in Databases

Metode yang digunakan dalam penelitian ini adalah Knowledge Discovery in Databases (KDD). KDD adalah proses penting yang mengidentifikasi pola – pola baru, valid, dan berpotensi berguna yang pada akhirnya dapat dipahami dalam data. Secara lebih detail, proses KDD seperti pada gambar berikut ini yang diadopsi dari[2].



Gambar 2.3 Knowledge Discovery in Databases

1. Selection

yang tujuan utamanya adalah untuk membuat kumpulan data target dari data asli, yaitu memilih subset variabel atau sampel data, di mana penemuan harus dilakukan

2. Pre-processing

yang bertujuan untuk membersihkan data dengan melakukan berbagai operasi, seperti pemodelan dan penghilangan noise, menentukan strategi yang tepat untuk menangani bidang data yang hilang, memperhitungkan informasi urutan waktu

3. Transformation

Merubah data sesuai format ekstention yang sesuai dalam pengolahan data mining karena beberapa metode pada data mining memerlukan format khusus sebelum dapat diproses pada data mining.

4. Data Mining

Proses utama pada metode yang diterapkan untuk mendapatkan pengetahuan baru dari data yang diproses. Pada penelitian ini diterapkan teknik clustering yaitu metode K-Means Clustering.

5. Interpretation

Mengidentifikasi pola – pola yang menarik kedalam knowledge base yang diidentifikasi. Pada tahap ini, menghasilkan pola – pola khas maupun model prediksi yang dievaluasi untuk menilai kajian yang ada sudah memenuhi target yang diinginkan.

2.5.4. Data Mining

Data mining merupakan metode pengolahan data berskala besar oleh karena itu data mining ini memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi. Secara umum kajian data mining membahas metode-metode seperti, clustering, klasifikasi, regresi, seleksi variable, dan market basket analisis[7]. Data mining adalah suatu proses ekstraksi atau penggalian data dan informasi yang sangat besar yang belum diketahui sebelumnya, namun dapat dipahami dari database yang besar dan digunakan untuk membuat keputusan bisnis penting. Proses ini memungkinkan pemasar menemukan pola-pola yang tidak diketahui dalam data yang telah dikumpulkan.

Secara umum ada dua jenis metode pada data mining, metode predictive dan metode descriptive. Metode predictive adalah proses untuk menemukan pola dari data yang menggunakan beberapa variable untuk memprediksi variabel lain yang tidak diketahui jenis atau nilainya. Teknik yang termasuk dalam predictive mining antara lain Klasifikasi, Regresi dan Deviasi. Metode descriptive adalah proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data. Teknik data mining yang termasuk dalam descriptive mining adalah Clustering, Association dan Sequential Mining [8]

2.5.5. Clustering

Clustering adalah proses mengelompokkan objek yang mirip ke dalam kelompok yang berbeda, atau lebih tepatnya membagi sebuah set data menjadi subset-subset yang memiliki makna yang sama untuk setiap subset. Setiap kelompok, atau cluster, terdiri dari objek yang memiliki kesamaan satu sama lain dan berbeda dari objek di kelompok lain [9]. Clustering juga

disebut segmentasi data dalam beberapa aplikasi karena melibatkan pembagian kumpulan data besar menjadi kelompok-kelompok berdasarkan kesamaan. Dalam machine learning, clustering adalah salah satu contoh dari unsupervised learning[10].

Clustering dapat digambarkan suatu proses data yang diorganisir dan dibagi menjadi beberapa kelompok sehingga membentuk pola serupa yang dikelompokkan bersama dalam cluster yang sama, sementara pola yang berbeda dikelompokkan dalam cluster yang berbeda[11]. Mengklasifikasikan atau mengelompokkan data ke dalam seperangkat kategori atau cluster adalah salah satu cara yang paling penting dalam hubungan dengan data, karena clustering dapat memainkan peran penting dalam kehidupan sehari-hari. Banyak aplikasi saat ini menggunakan clustering di berbagai bidang. Contoh pengelompokan data yang digunakan untuk analisis data statistik adalah pengelompokan untuk pembelajaran mesin, data mining, pengenalan pola, analisis gambar, dan bioinformatika.

2.5.6. Metode Clustering

Pada dasarnya clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain. Clustering merupakan salah satu metode data mining yang bersifat tanpa arahan (unsupervised), maksudnya metode ini diterapkan tanpa adanya latihan (training) dan tanpa ada guru (teacher) serta tidak memerlukan target output. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu hierarchical clustering dan non-hierarchical clustering[7]

Hierarchical clustering adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga cluster akan membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas

antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah cluster. Dendrogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut[12]. Berbeda dengan metode hierarchical clustering, metode non-hierarchical clustering justru dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan (dua cluster, tiga cluster, atau lain sebagainya). Setelah jumlah cluster diketahui, baru proses cluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan K-Means Clustering[12].

2.5.7. K-Means

Algoritma *K-Means* adalah salah satu algoritma partitional yang menggunakan nilai awal *centroid* untuk menghitung jumlah awal kelompok[13]. Untuk membentuk kelompok data, algoritma *K-Means* menggunakan proses berulang. Jumlah *cluster* awal yang diinginkan dimasukkan sebagai *input*, dan algoritma ini menghasilkan titik *centroid* akhir sebagai *output*. Metode *K-Means* akan memilih titik awal *centroid* secara acak. Jumlah iterasi yang diperlukan untuk mencapai *centroid cluster* dipengaruhi oleh inisialisasi awal yang acak.

Dasar algoritma *K-Means* sebagai berikut:

1. Tentukan nilai k sebagai jumlah *cluster* yang ingin dibentuk
2. Inisialisasi k sebagai *centroid* yang dapat dibangkitkan secara random
3. Hitung jarak setiap data ke masing-masing *centroid*
4. menggunakan persamaan *Euclidean Distance* yaitu sebagai berikut:

$$d(P, Q) = \sqrt{\sum_{j=1}^p (X_j(P) - X_j(Q))^2}$$

5. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*-nya.
6. Tentukan *centroid* baru k
7. Kembali ke Langkah 3 jika posisi *centroid* baru dan *centroid* lama tidak sama.

2.5.8. *K-Means++*

Pada metode K-Means, pemilihan centroid dilakukan secara acak yang merupakan kelemahan dari proses clustering K-Means. Untuk mengatasi kelemahan tersebut dikembangkan metode K-Means++. Metode ini melakukan inisialisasi centroid dengan lebih optimal dan meningkatkan kualitas clustering. Kedua metode tersebut, K-Means dan K-Means++, hanya berbeda pada pilihan pusat/centroid awal. K-Means memilih pusat cluster secara acak sesuai dengan nilai K yang ditentukan, sedangkan K-Means++ secara acak memilih satu centroid dan kemudian menghitung jarak centroid dari data lainnya. Nilai maksimum berarti jarak terjauh dan peluang tertinggi untuk menjadi centroid baru[14].

Langkah – langkah K-Means++ sebagai berikut:

1. Menentukan titik centroid pertama secara acak
2. Menghitung jarak semua titik dalam dataset dari centroid yang dipilih. Jarak titik x_i dari centroid terjauh dapat dihitung menggunakan formula dibawah ini.

$$P(i) = \frac{d_{(i)}}{\sum_{x^1 \in x} d(i)}$$

Keterangan : $d_{(i)}$ = Jarak Euclidean Distance

$\sum_{x^1 \in x} d(i)$ = Jumlah Jarak Euclidean Distance

3. Menjadikan titik P_i sebagai centroid baru yang memiliki probabilitas maksimum. Formula probabilitas seperti di bawah ini.
4. Ulangi langkah 2-3 sampai menemukan K-Centroid.
5. Setelah mendapatkan centroid, langkah selanjutnya sama dengan metode K-Means.

2.5.9. *Euclidean Distance*

Euclidean Distance merupakan metode perhitungan jarak antar dua titik dalam ruang yang diperkenalkan oleh matematikawan asal Yunani bernama Euclid. Dalam bahasa matematika euclidean distance seperti

perhitungan Pythagoras yang digunakan untuk mengukur dua titik dalam satu dimensi[15].

Berikut adalah rumus perhitungan *Euclidean Distance*:

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

Euclidean Distance memiliki kelebihan dan kekurangan dibandingkan metode perhitungan lain. Kelebihan perhitungan *Euclidean Distance* sebagai berikut[16]:

1. Perhitungan jarak menggunakan *Euclidean Distance* lebih umum digunakan dalam metode *K-Means*.
2. *Euclidean Distance* mempunyai hasil yang lebih optimal dibandingkan dengan perhitungan yang lainnya.
3. Referensi dari *Euclidean Distance* lebih banyak karena secara umum perhitungan *Euclidean Distance* digunakan dalam perhitungan jarak pada *K-Means Clustering*.

Euclidean Distance juga memiliki kekurangan sebagai berikut:

1. Perhitungan manual menggunakan *Euclidean Distance* cenderung lebih lama karena jumlah iterasi *Euclidean Distance* lebih banyak dibandingkan dengan *Manhattan Distance*.

2.5.10. Davies-Bouldin Index

Dalam analisis data, indeks Davies-Bouldin digunakan untuk mengukur kualitas *clustering*. Tujuan indeks ini adalah untuk mengukur seberapa baik *cluster* yang dibuat oleh algoritma *clustering* memisahkan kelompok data yang berbeda dan mendekati pusat *cluster-nya*; nilai indeks yang lebih rendah menunjukkan bahwa *clustering* yang dihasilkan lebih baik[17].

Rumus *Davies-Bouldin Index* untuk menghitung kualitas *clustering* antara dua *cluster* C_i dan C_j adalah sebagai berikut:

$$R_{ij} = \frac{S_i + S_j}{D_{ij}}$$

Di mana:

1. R_{ij} adalah nilai *Davies-Bouldin Index* antara *cluster* C_i dan C_j .
2. S_i adalah ukuran dispersi atau sebaran data dalam *cluster* C_i , yang dapat dihitung dengan variasi atau jarak rata-rata antara titik data dalam klaster dengan pusat *cluster* m_i .
3. S_j adalah ukuran dispersi atau sebaran data dalam *cluster* C_j , yang dapat dihitung dengan variasi atau jarak rata-rata antara titik data dalam *cluster* dengan pusat *cluster* m_i .
4. D_{ij} adalah jarak antara pusat *cluster* m_i dan m_j

Langkah-langkah umum untuk menghitung *Davies-Bouldin Index* adalah sebagai berikut:

1. Hitung pusat *cluster* m_i dan m_j untuk masing-masing *cluster* C_i dan C_j .
2. Hitung ukuran dispersi S_i untuk *cluster* C_i dan S_j untuk *cluster* C_j . Ini dapat dilakukan dengan menghitung jarak rata-rata antara titik-titik data dalam *cluster* dengan pusat *cluster*.
3. Hitung jarak d_{ij} antara pusat *cluster* m_i dan m_j .
4. Gunakan rumus *Davies-Bouldin Index* untuk menghitung nilai R_{ij} antara *cluster* C_i dan C_j .
5. Ulangi Langkah 1 hingga 4 untuk setiap pasangan *cluster* yang mungkin.

Setelah semua nilai R_{ij} dihitung, *Davies-Bouldin Index* keseluruhan dapat dihitung dengan rumus:

$$DBI = \frac{1}{N} \sum_{i=1}^N \max_{j \neq i} R_{ij}$$

Di mana:

1. DBI adalah Davies-Bouldin Index keseluruhan.
2. N adalah jumlah total *cluster*.

Nilai *Davies-Bouldin Index* yang lebih rendah menunjukkan kualitas *Clustering* yang lebih baik, karena menunjukkan bahwa *cluster-nya* lebih terpisah dan lebih dekat dengan pusat *cluster-nya*.

2.5.11. Unified Modelling Language

Unified Modelling Language (UML) adalah Salah satu standar industri untuk visualisasi, desain, dan dokumentasi sistem piranti lunak. UML menyediakan standar untuk merancang model sistem[18]. Dengan menggunakan UML kita dapat membuat model untuk semua jenis aplikasi perangkat lunak, yang dapat berjalan pada perangkat keras, sistem operasi, atau jaringan apa pun, dan dapat ditulis dalam bahasa pemrograman apa pun.

2.5.12. Use Case Diagram

Use case diagram merupakan pemodelan untuk kelakuan (behavior) sistem informasi yang akan dibuat. Use case mendeskripsikan sebuah interaksi antar satu atau lebih aktor dengan sistem informasi yang akan di buat[19].

2.5.13. Class Diagram

Class adalah sebuah spesifikasi yang jika diinstansiasi akan menghasilkan sebuah objek dan merupakan inti dari pengembangan dan desain berorientasi objek. *Class* menggambarkan keadaan atribut atau properti suatu sistem, sekaligus menawarkan layanan untuk memanipulasi keadaan tersebut. *Class diagram* menggambarkan struktur dan deskripsi *class*, *package* dan objek beserta hubungan satu sama lain seperti containment, pewarisan, asosiasi, dan lain-lain[18].

2.5.14. Activity Diagram

Activity Diagram mewakili berbagai aliran aktivitas dalam sistem yang Anda rancang, bagaimana setiap aliran dimulai, keputusan apa yang dibuat, dan bagaimana berakhir. *Activity Diagram* juga dapat mewakili proses paralel yang dapat terjadi pada beberapa eksekusi[18].

2.5.15. Sequence Diagram

Diagram Sequence menggambarkan perilaku objek dalam use case dengan menggambarkan masa hidup mereka dan pesan yang dikirim dan diterima di antara mereka[19].

2.5.16. Python

Python adalah bahasa pemrograman yang dikembangkan oleh Guido van Rossum dan populer sebagai bahasa pemrograman Web serta bahasa skripting. Menurut pengertiannya, Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode[20]. Python digambarkan sebagai bahasa yang memiliki banyak kemampuan dan kemampuan bersama dengan sintaksis kode yang sangat jelas. Selain itu, memiliki banyak fitur pustaka standar yang luas dan komprehensif. Python sebagian besar mendukung berbagai paradigma pemrograman, tetapi tidak terbatas pada pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional.

2.5.17. Website

Situs web, juga disebut sebagai "*website*", adalah halaman web yang saling berhubungan yang biasanya terletak pada peladen yang sama dan berisi kumpulan data yang disediakan oleh individu, kelompok, atau organisasi. Sebuah situs web biasanya terdiri dari setidaknya satu server web yang dapat diakses melalui jaringan seperti Internet atau jaringan wilayah lokal (LAN). Alamat Internet yang diberikan oleh server web ini dikenal sebagai URL[20].