

BAB II

LANDASAN TEORI

2.1 Kajian Terhadap Penelitian Terkait Sebelumnya

Analisis terhadap penelitian terkait yang berfungsi sebagai referensi untuk penelitian ini:

2.1.1 Analisis perbandingan algoritma Naïve Bayes, k-Nearest Neighbor dan Neural Network untuk permasalahan class-imbalanced data pada kasus credit card fraud dataset.

Studi sebelumnya menunjukkan bahwa tingginya keinginan orang untuk melakukan transaksi dengan kartu kredit di sektor perbankan memiliki potensi meningkatkan tindakan penipuan kartu kredit. Dataset mengenai tindakan penipuan kartu kredit memiliki 284.807 data yang dikumpulkan melalui Kaggle digunakan dalam penelitian. Dataset ini memiliki kelas yang tidak seimbang, dengan perbandingan antara kelas sebesar 99,8% banding 0,2%. Penelitian ini berharap untuk menyelesaikan masalah kelas tidak seimbang ini dengan menggunakan sampling bawah. Perbandingan antara algoritma Naive Bayes, KNN, dan *Neural Network* akan dilakukan untuk mengetahui klasifikasi mana yang paling cocok untuk mengatasi masalah data yang dibalanced kelas. Dalam penelitian ini, T-test dibuat untuk menentukan apakah perbedaan antar algoritma signifikan. Nilai akurasi dan area di bawah kurva (AUC) digunakan untuk menilai kinerja algoritma. Karena memiliki nilai akurasi tertinggi mencapai 93,59% dan nilai AUC mencapai 0,977, Neural Network menunjukkan kinerja yang lebih baik dari algoritma lainnya.

Menurut hasil uji, Neural Network dengan KNN menunjukkan perbedaan yang lebih besar dibandingkan dengan Neural Network dengan Naïve Bayes.

2.1.2 Perbandingan Akurasi Algoritma Random Forest dan Algoritma Multi Layer Perceptron untuk Klasifikasi Penyakit Diabetes.

Studi sebelumnya oleh Eky Cahya Putra Witjaksana, Rd. Rohmat Saedudin, dan Vanda Pradwiyasmana Widartha mengenai kadar gula darah yang tinggi dan masalah metabolisme untuk karbohidrat, lipid, dan protein adalah tanda penyakit diabetes yang berlangsung lama. Mengganggu metabolisme ini disebabkan oleh ketidakmampuan insulin untuk berfungsi dengan benar. Indonesia berada di peringkat ke-7 dari sepuluh negara dengan pasien diabetes terbanyak di dunia, menunjukkan bahwa orang harus waspada terhadap diabetes, menurut International Diabetes Federation (IDF). Dipenelitian ini, penulis menggunakan Machine Learning untuk mendapatkan prediksi untuk penyakit diabetes; dan dataset untuk penelitian ini adalah pima-indian diabetes yang memiliki 768 data, yang mencakup delapan atribut data : kehamilan, glukosa, tekanan darah, ketebalan kulit, insulin, BMI (indeks massa tubuh), fungsi genealogi diabetes, usia, dan hasil. Untuk melakukan klasifikasi dataset pima-indian-diabetes, penelitian ini akan melakukan perbandingan hasil akurasi dari Random Forest dan Multilayer Perceptron. Sebelum melakukan perbandingan, penulis membersihkan dataset, membuat fitur matriks , membagi data, dan meningkatkan fitur. Setelah data diproses dalam tahap preprocessing, langkah selanjutnya adalah menemukan akurasi terbaik dengan menggunakan Confusion Matrix untuk mengetahui ROC AUC dan F1-Score dari setiap algoritma yang digunakan.

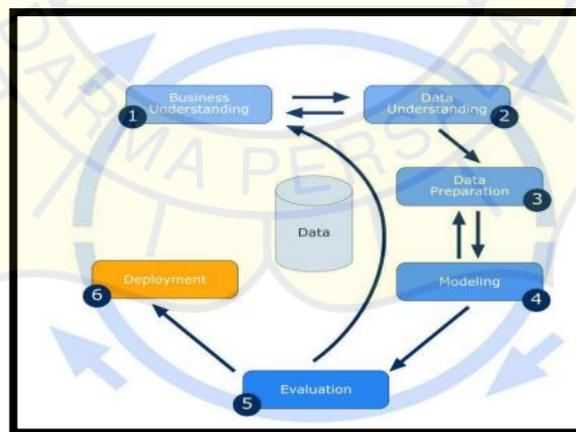
2.1.3 Algoritma Klasifikasi Decision Tree Untuk Rekomendasi Buku Berdasarkan Kategori Buku.

Penelitian sebelumnya menyatakan bahwa dengan semakin berkembangnya teknologi, semakin beragamnya buku yang tersedia di internet. sebagai sistem rekomendasi yang digunakan di situs web buku daring yang memiliki buku yang relevan dan sesuai dengan kesukaan seseorang. GoodReads adalah situs jejaring sosial yang fokus pada katalog buku dan memungkinkan pengguna berbagi buku bacaan rekomendasi satu sama lain melalui peringkat, penilaian, dan komentar. Ada banyak data yang dapat diproses dengan metode pembelajaran mesin dalam buku besar situs rekomendasi, namun masih belum dikenal sebagai model yang paling akurat. Kami dapat memberikan rekomendasi yang lebih akurat dengan menggunakan model yang tepat. Oleh karena itu, penelitian ini akan menyelidiki data yang dikumpulkan dari kumpulan data goodreads-books dan www.kaggle.com. Studi ini mengusulkan klasifikasi model data mining untuk menentukan model yang paling cocok untuk rekomendasi buku GoodReads. Decision Tree, KNN, Nave Bayes, Random Forest, dan SVC (*Support Vector Classifier*) adalah algoritma yang digunakan untuk transmisi model menggunakan akurasi, presisi, memori, skor f1, gangguan matriks, dan AUC. Hasil uji dari beberapa klasifikasi klasifikasi menunjukkan bahwa Pohon Keputusan memiliki akurasi mencapai 99,95%, presisi mendapatkan 100%, recall mencapai 96%, skor f1 mendapatkan 98% dengan MAE sebesar 0,05.

2.2 Cross-Industri Standard Process for Data Mining

CRISP-DM merupakan salah satu metodologi yang digunakan dalam proses penambangan data, yang didirikan oleh lima perusahaan pada tahun 1996. Metodologi ini kemudian dikembangkan oleh ratusan organisasi dan perusahaan di Eropa untuk dijadikan metode standar non-proprietary untuk data mining. Setelah itu, ratusan perusahaan dan lembaga di seluruh Eropa mengembangkan metode standar non-kepemilikan untuk penambangan data. Versi pertama dari metodologi ini dipresentasikan di Brussels pada Maret 1999 pada Lokakarya CRISP-DM SIG ke-4 (Pete Chapman, 1999). Pada tahun berikutnya, proses data mining yang didasarkan pada model ini dipublikasikan (Pete Chapman, 2000).

Pada penelitian ini, sebagai pemecah masalah penulis akan menggunakan metodologi data mining CRISP-DM. Proses metodologi ini terdiri dari enam fase, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*.



Gambar 2.1 Skema CRISP-DM

(Sumber : Colin Shearer 2012)

1. *Business Understanding*

Pada langkah ini dimulai dengan memahami tujuan bisnis dari proyek data mining.

Tujuan ini harus jelas dan spesifik sehingga semua langkah analisis data dapat ditunjukkan untuk mencapainya.

2. *Data Understanding*

Proses ini, data yang tersedia untuk proyek dieksplorasi dan dipahami lebih dalam. Ini meliputi pengumpulan data, deskripsi data, dan pemeriksaan awal untuk menentukan kualitas dan kecocokannya dengan tujuan proyek. Jika diinginkan, pilih sebagian kecil grup data yang mungkin mengandung pola dari permasalahan.

3. *Data Preparation*

Pada titik ini ini melibatkan persiapan data yang meliputi pembersihan data (data cleaning), transformasi data, integrasi data jika perlu, serta pemilihan subset data yang relevan untuk analisis lebih lanjut. Tujuan dari fase ini adalah mempersiapkan data yang bersih dan siap untuk pemodelan.

4. *Modeling*

Tahap ini, berbagai teknik pemodelan data mining diterapkan untuk menganalisis data. Ini termasuk pemilihan teknik pemodelan yang sesuai, pengaturan parameter, dan pengembangan model untuk memenuhi tujuan bisnis yang telah ditetapkan.

5. *Evaluation*

Setelah model dibuat, evaluasi dilakukan untuk mengukur kinerja dan keefektifan model. Evaluasi ini membantu memastikan bahwa model dapat dipercaya dan sesuai dengan tujuan bisnis yang ingin dicapai.

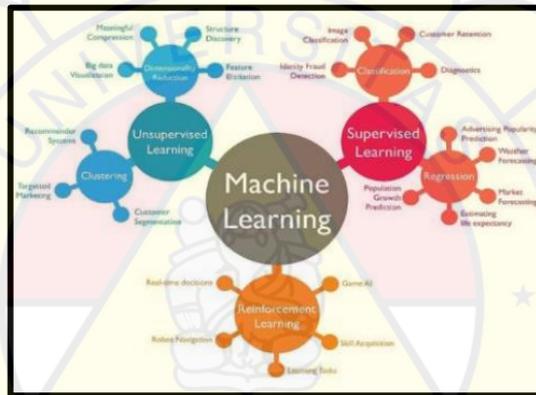
6. *Deployment*

Fase terakhir CRISP-DM melibatkan implementasi atau penyebaran solusi yang dihasilkan dari proses data mining ke lingkungan operasional. Hal ini meliputi

dokumentasi, integrasi dengan sistem yang ada, dan pelatihan pengguna jika diperlukan.

2.3 Machine Learning

Machine Learning (ML) atau biasa disebut juga sebagai Mesin Pembelajaran ialah cabang dari kecerdasan buatan yang fokus pada belajar dari data, atau belajar dari data. Fokusnya adalah membuat sistem yang dapat belajar secara “mandiri” tanpa diprogram ulang oleh manusia. Untuk mendapatkan hasil output yang optimal, ML membutuhkan informasi yang tepat sebagai bahan belajar (dalam proses pelatihan).

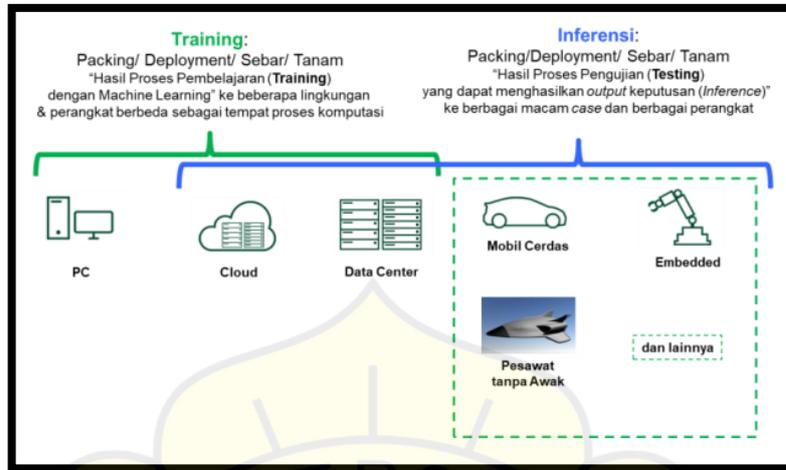


Gambar 2.2 Map Cangkupan dari Machine Learning

(Sumber : Buku AI, Machine Learning & Deep Learning)

Hasil pengembangan produk berbasis AI (machine learning dan teknologi , dll) hendaknya dapat lebih mudah dan lancar diimplementasikan di masyarakat setempat atau bahkan masuk kancah bisnis nasional maupun internasional . Berikut

adalah contoh peta produk yang dapat diterapkan pada masyarakat .daripeta produk yang dapat diterapkan ke masyarakat.



Gambar 2.3 Map Pengembang Produk App Learning
(Sumber : Buku AI, Machine Learning & Deep Learning)

Tipe Data Dalam Data mining

Data dapat dikategorikan menjadi kategorik dan numerik berdasarkan nilai atau isi mereka . Sangat penting untuk memahami kedua jenis data ini karena ukuran statistik atau analisis yang akan digunakan akan mempengaruhi metode data mining yang akan dipilih.

Kategori Data —Digunakan untuk data yang tidak dapat dihitung secara kuantitatif karena tidak dapat menggunakan operasi matematika seperti penjumlahan dan perkalian. Nilainya mungkin tidak sama. Data nominal, biner, dan ordinal adalah komponen dari data kategorik.

Data Nominal

Data nominal adalah jenis data dasar yang nilainya hanya berupa label atau nama dan tidak dapat diurutkan atau diukur. Contoh data nominal termasuk nama,

kategori, teks, tanggal, waktu, dan mata uang, tetapi kita dapat menghitung frekuensinya. Oleh karena itu, untuk data ini, ukuran statistik yang digunakan adalah ruang.

Data Biner

Data biner adalah representasi dari informasi menggunakan sistem angka biner, yang terdiri dari dua angka yaitu 0 dan 1. Setiap angka biner mewakili satu bit (binary digit), dan serangkaian bit-bit ini digunakan untuk merepresentasikan data yang lebih kompleks. Di dalam komputer dan teknologi digital lainnya, data biner digunakan sebagai dasar untuk menyimpan, memproses, dan mentransmisikan informasi.

Dalam konteks komputasi, data biner sering kali diorganisir dalam bentuk byte (delapan bit), dan dari sini dapat diinterpretasikan sebagai berbagai jenis data, seperti angka, teks, gambar, atau suara. Sistem biner sangat penting karena komputer menggunakan sirkuit elektronik yang memanfaatkan keadaan "on" dan "off" untuk melakukan operasi logika, penghitungan, dan penyimpanan data.

Data Ordinal

Data ordinal adalah kumpulan data yang dapat diurutkan dari tingkat paling rendah ke tingkat tertinggi. Tetapi jarak antara tingkatan tidak dapat diukur atau dibandingkan dengan angka. Nilai tengah data dapat ditemukan dengan menggunakan ukuran statistik median selain modus. Data numerik—digunakan untuk data yang dapat diukur secara kuantitatif dan memungkinkan operasi

matematik dilakukan. Rasio data dan interval termasuk dalam kategori data numerik.

Data Interval

Untuk memungkinkan operasi penjumlahan dan pengurangan, interval data sudah memiliki jarak antar level. Karena jarak data tidak disetel ke nol, perkalian dan pembagian tidak dapat dilakukan. Untuk data ini, ukuran statistik mean atau median, serta modus dan median, dapat digunakan.

Data Rasio

Rasio data adalah data numerik absolut yang bernilai nol. Keunggulan data rasio adalah mereka dapat melakukan semua operasi matematika, seperti menjumlahkan, mengurangi, perkalian, dan membagi.

2.4 Properti

Menurut kamus besar besar Indonesia properti adalah tanah serta bangunan dan sarana-prasarana yang merupakan satu kesatuan dengan tanah dan/atau bangunan yang bersangkutan.

Mengutip Buku Hukum Properti yang ditulis Dhaniswara K. Harjono, properti adalah suatu bangunan (bertingkat) mendatar atau vertikal yang digunakan sebagai bangunan tempat tinggal (residence), atau bangunan untuk tempat komersial dan tempat lainnya (bukan tempat tinggal).

Properti adalah suatu harta benda, harta benda, atau tanah yang dapat bergerak maupun tidak bergerak. Istilah properti juga mengacu pada sesuatu yang

secara umum dikenal sebagai suatu kesatuan yang berada di bawah kepemilikan atau hak eksklusif seseorang atau sekelompok orang. Bentuk utama properti adalah real estate (tanah), properti pribadi (properti fisik barang), dan kekayaan intelektual.

2.5 Implementasi

Menurut Mulyadi (2018:12), Implementasi berarti melakukan aksi untuk mencapai tujuan (*goals*) yang telah diputuskan. Tujuan dari kegiatan ini adalah mentransformasikan keputusan-keputusan tersebut ke dalam pola-pola tindakan dan untuk mendapatkan perubahan besar atau kecil dari cara yang telah ditetapkan sebelumnya. Implementasi pada dasarnya mencoba memahami yang seharusnya terjadi setelah sistem dijalankan.

2.6 Aplikasi

Menurut (Juansyah & Refianti, 2018), “Aplikasi adalah penggunaan semenjak rancang susunan menjelang membenahi petunjuk yang mengabdikan hukum atau doktrin tekanan perencanaan tertentu.” Dengan adanya aplikasi, pemakai bisa mengamalkan berbagai bagian, seumpama membenahi petunjuk, menyebrangi hiburan, belajar, dan mengamalkan bagian-bagian lainnya.

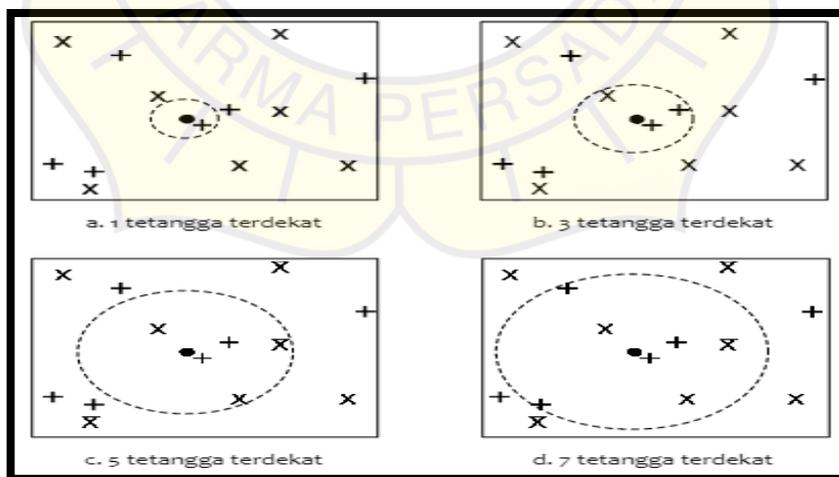
Aplikasi adalah suatu program dalam perangkat komputasi yang dirancang untuk menjalankan suatu tugas atau fungsi tertentu kepada pengguna. Secara umum, aplikasi digunakan untuk memberikan solusi atau memfasilitasi kegiatan pengguna dalam melakukan pekerjaan, mengakses informasi, bermain game, berkomunikasi, atau melakukan aktivitas lainnya dengan bantuan komputer atau perangkat digital lainnya.

2.7 Metode K-Nearest Neighbors

Menurut Eko Prasetyo (2014, 150) Metode K-Nearest Neighbor menjadi

salah satu pendekatan berbasis Neural Network yang paling lama dan populer. Nilai K yang dipakai di penelitian ini menunjukkan banyaknya objek/tetangga terdekat yang terlibat dalam menentukan hasil prediksi label kelas dari data eksperimen. Di antara K tetangga dekat yang dipilih, pemungutan suara kelas dilakukan terhadap K tetangga dekat. (Tan et al, 2005).

Pada Gambar 2.4-(a) digunakan 1 tetangga terdekat. Tetangga tersebut mempunyai label "+" maka diprediksi data uji tersebut masuk ke kelas "+" Pada Gambar 2.4 (b) digunakan 3 tetangga terdekat. kelas "+" ada 2 suara, sedangkan kelas x ada 1 suara. Karena jumlah suara kelas "+" lebih banyak daripada kelas "x", data uji tersebut diprediksi masuk ke kelas "+". Pada Gambar 2.4 (c) digunakan 5 tetangga terdekat, kelas ada 2 suara, sedangkan kelas "X" ada 3 suara. Karena jumlah suara kelas "x" lebih banyak daripada kelas data uji tersebut diprediksi masuk ke kelas "x". Demikian juga Gambar 2.4(d) yang menggunakan 7 tetangga terdekat, data uji diprediksi masuk ke kelas "+" karena karena jumlah suara kelas lebih banyak daripada kelas.



Gambar 2.4 KNN dengan Nilai K Tetangga

2.8 Multi Layer Perceptron

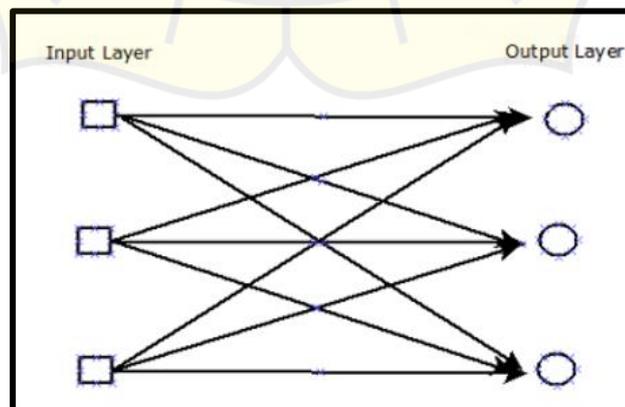
Jaringan syaraf tiruan lapis banyak atau Multi Layer Perceptron merupakan

lanjutan dari jaringan saraf tiruan lapis banyak, juga dikenal sebagai perceptron lapis banyak. Dalam algoritma delta yang disebut error backpropagation training, argumen yang dimasukkan diumpamakan secara arah maju, sedangkan proses pembelajaran juga menggunakan perambatan arah balik selain perambatan arah maju. Apabila hasil tidak sesuai dengan harapan, bobot akan diperbarui sepanjang siklus pembelajaran hingga nilai kemelesetan minimum yang diharapkan tercapai atau hasil sama dengan harapan (Muis, 2006).

Kemampuan untuk data pelatihan adalah keunggulan utama dari sistem Multi Layer Perceptron (MLP). Kelemahan utama dari MLP adalah mereka membutuhkan pelatihan untuk beroperasi dan membutuhkan waktu yang lama untuk memprosesnya .

Berdasarkan jumlah Layer arsitektur jaringan *Multi Layer Perceptron* (MLP) dapat di klasifikasikan menjadi dua kelas yang berbeda yaitu:

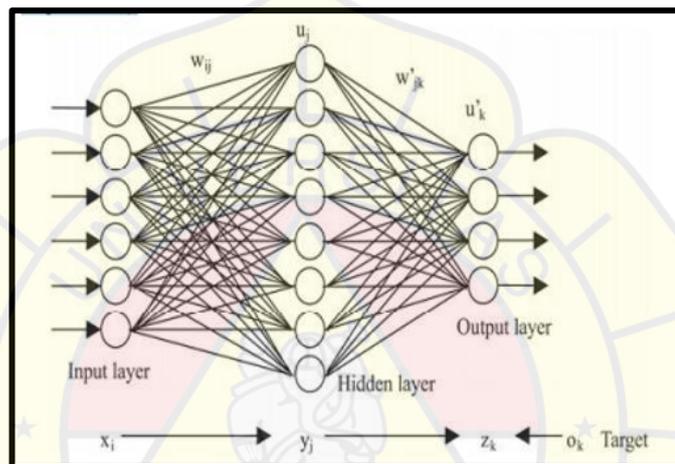
1. Jaringan layer tunggal (*single layer network*) merupakan semua unit *input* dalam jaringan ini dihubungkan dalam semua unit *output*, meskipun dengan bobot yang berbeda – beda. Berikut contoh jaringan single layer dapat di lihat pada gambar 2.5



Gambar 2.5 Jaringan *Single Layer Network*

(Sumber: Kusumadewi, 2013)

2. Jaringan layar jamak (*multi layer network*) biasanya mengacu pada jaringan yang terdiri dari beberapa lapisan neuron atau unit pemrosesan informasi yang saling terhubung. Contoh paling umum adalah jaringan saraf tiruan feedforward multilayer perceptron (MLP), yang terdiri dari setidaknya satu lapisan input, satu atau lebih lapisan tersembunyi (*hidden layer*), dan satu lapisan output. Contoh jaringan layar jamak dapat dilihat pada gambar 2.5 (Sutojo,2014).



Gambar 2.6 Jaringan *Multi Layer Network*

(Sumber: Widodo, 2013)

2.9 Data Mining

2.9.1 Pengertian Data Mining dan Bukan Data Mining

Menurut Eko Eko Prasetyo dalam DATA MINING Mengolah Data Menjadi Informasi Menggunakan Matlab (2014, 85) Pada tahun 1990, istilah data mining mulai digunakan. Seiring berjalannya waktu, penggunaan informasi menjadi penting dalam berbagai bidang, seperti akademisi, bisnis, dan kesehatan (Gorunescu, 2011). Data mining sangat berguna di banyak bidang yang memiliki

banyak data. Namun, karena bidang penelitian ini relatif baru dan belum melewati masa remajanya, data mining masih membingungkan tentang keahliannya.

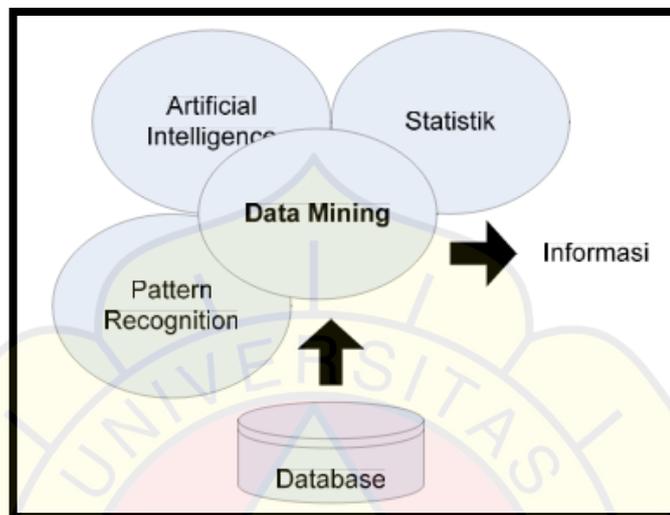


Gambar 2.7 Data Mining

Meskipun data mining masih sangat baru, diproyeksikan menjadi jutaan dolar di industri pada tahun 2000, tetapi sejumlah peneliti melihatnya sebagai kata yang kotor dalam statistik (Gorunescu, 2011). Pada saat itu, mereka tidak menyukai data mining.

Munculnya data mining didasari oleh semakin banyaknya informasi yang disimpan dalam database. Misalnya di supermarket, berapa banyak transaksi pelanggan yang dilakukan per hari dan berapa juta data yang disimpan per bulan, bahkan tahunan. Contoh lain adalah Anda mungkin sering menerima surat penawaran barang atau jasa jika Anda memiliki kartu kredit. Bank harus mengeluarkan 500 juta rupiah jika memiliki 1 juta nasabah dan biaya untuk pengiriman surat setiap nasabah adalah 500 rupiah. Namun, hanya 15 persen nasabah yang benar-benar membeli, sehingga 85 persen dari 500 juta, atau 425 juta, dibuang sia-sia. Biaya pengiriman dapat dikurangi jika bisnis dapat menggunakan data yang ada sehingga hanya pelanggan yang berpotensi membeli yang dikirim surat. Data perusahaan semakin banyak, dan pertanyaannya adalah apakah akan

digunakan untuk apa? Apakah hanya untuk dibuat laporan akhir tahun dan kemudian dibuang? Apakah hanya akan disimpan di gudang data tanpa manfaat apa pun? Tentu saja akan sangat menyedot jika data ini tidak digunakan untuk kepentingan perusahaan atau lembaga yang berkepentingan.



Gambar 2.8 Akar Ilmu Data Mining

Istilah lain yang memiliki arti sama dengan data mining, ialah Data Mining Knowledge-discovery (KDD). Dikenal juga sebagai KDD, bertujuan untuk menggunakan database dengan mengolahnya untuk mendapatkan informasi baru yang bermanfaat. Bidang ilmu data mining memiliki empat akar keilmuan, seperti yang ditunjukkan pada Gambar 2.8, jika dilacak akar keilmuannya, ternyata data mining mempunyai empat akar bidang ilmu sebagai berikut :

1. Statistik

Bidang ini memiliki sejarah yang sangat tua, jadi tanpa statistik, data tambang mungkin tidak ada. Ketika variabel sifat tidak memiliki informasi alami yang cukup, EDA berguna untuk mengidentifikasi hubungan sistematis antar mereka.

Salah satu teknik EDA konvensional yang digunakan dalam penambangan data adalah

- **Metode Komputasional** : statistik deskriptif seperti distribusi, tabel frekuensi, korelasi, metode penelitian multivariat seperti analisis kelompok, faktor , analisis dan klasifikasi komponen utama, analisis kanonik, analisis diskriminan, klasifikasi pohon , dan analisis kecerahan; model linier non-linier tingkat lanjut seperti regresi linier non-linier, perkiraan penurunan waktu, dll.
- **Visualisasi Data** : menghasilkan penyajian data yang lebih visual, yang dianggap paling bermanfaat. Pada saat yang sama, visualisasi data adalah teknik eksplorasi data yang menarik. Semua jenis histogram, seperti lingkaran, batang, silinder, kerucut, piramida, dan sebagainya, adalah teknik visualisasi yang paling terkenal.

2. Kecerdasan Buatan alias Artificial Intelligence (AI)

Bidang studi ini tidak sama dengan statistik. Sebagai hasil dari teorinya yang didasarkan pada metode heuristik, kecerdasan buatan terlibat dalam metode pemrosesan informasi yang berbasis pada model penalaran manusia. Penambangan data menunjukkan bahwa sistem komputer belajar sambil melakukan sesuatu. Pembelajaran mesin adalah cabang penting dari kecerdasan buatan.

3. Pengenalan Pola

Sebenarnya, data mining adalah hasil dari pengenalan pola, tetapi hanya berkaitan dengan data dari database. Informasi yang diolah dari database tidak dalam bentuk relasional, tetapi dalam bentuk normal pertama, yang merupakan

bentuk dataset yang dibentuk. Namun, pola asosiasi pencarian dan pola urutan adalah ciri khas data mining.

4. Sistem Basis Data

dasar bidang ilmu keempat yang dikenal sebagai data mining, yang menghasilkan informasi dari data yang telah diekstraksi melalui teknik yang telah disebutkan sebelumnya.

Meskipun data mining juga berarti mencari data, tidak semua pengambilan data/informasi disebut data mining. Berikut ini adalah beberapa contoh apa yang disebut dengan :

- **Bukan data mining** : mencari informasi spesifik di Internet (misalnya, mencari informasi tentang memasak dengan Google).

Data mining: Mengelompokkan informasi serupa dalam hasil penelusuran ke dalam konteks tertentu (misalnya makanan Indonesia, makanan Belanda, dll. yang ditemukan Google).

- **Bukan data mining** : Dokter mencari rekam medis untuk menganalisis informasi tentang pasien dengan penyakit tertentu.

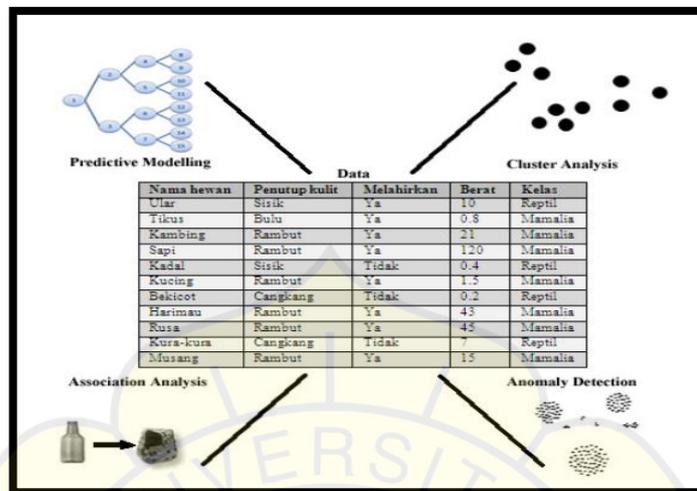
Data mining : Peneliti medis menemukan cara untuk mengelompokkan rekam medis pasien berdasarkan informasi diagnostik, usia, dan alamat.

- **Bukan data mining** : Analisis Visual Laporan Keuangan Penjualan Perusahaan.

Data mining : Menggunakan database peristiwa bisnis yang berfokus pada data penjualan untuk mengidentifikasi profil pelanggan utama.

2.9.2 Pekerjaan Dalam Data Mining

Pekerjaan penambangan data dapat dibagi menjadi pemodelan prediktif, analisis kluster, analisis asosiasi, dan deteksi anomali



Gambar 2.9 Pekerjaan Utama Data mining

1. Model Prediksi (prediction modelling)

Pekerjaan ini mencakup pembuatan model yang memiliki kemampuan untuk memetakan setiap kumpulan variabel ke masing-masing target, dan kemudian menggunakan model ini untuk menetapkan nilai target ke kumpulan hasil baru. Dua jenis model prediktif adalah regresi dan klasifikasi. Regresi digunakan untuk variabel target kontinu, sedangkan klasifikasi digunakan untuk variabel target diskrit.

2. Analisis cluster (cluster analysis)

Analisis massal mengelompokkan data ke dalam beberapa kelompok berdasarkan kemiripan karakteristik masing-masing data pada kelompok yang ada. Data dalam batas kesamaan suatu kelompok diasosiasikan dengan kelompok tersebut dan ditugaskan ke kelompok lain ketika data tersebut berada di luar batas kesamaan kelompok.

3. Analisis asosiasi (association analysis)

Analisis asosiasi digunakan untuk menemukan model yang menunjukkan kekuatan hubungan antar fitur dalam data; pola yang ditemukan biasanya termasuk dalam aturan atau fitur implisit. Tujuannya adalah menemukan model yang menarik secara efektif.

2.10 Bahasa Pemrograman

Bahasa pemrograman adalah sekumpulan aturan sintaksis dan semantik yang memungkinkan pembuat program menginstruksikan komputer untuk melakukan tugas tertentu, seperti menghitung angka, memproses teks, atau mengontrol perangkat keras, serta membuat program aplikasi (Azis & Sarmidi, 2018) (Eka Achyani & Saumi, 2019).

2.10.1 HTML (Hypertext Markup Language)

Menurut Muhamad Hendra Sunarya dan Muhamad Bahid (2020, 1) HTML ialah bahasa markup yang dipakai untuk membuat halaman web yang berisi berbagai data yang dapat diakses melalui Internet (Wempen, 2011). Seperti facebook.com., Twitter.com, google.com ditampilkan dalam HTML Jadi bisa dibidang HTML adalah bahasa dasar web browser untuk menampilkan halaman web (Ariona, 2013).

Hypertext, markup, dan bahasa membentuk HTML. Menurut Pratama (2016), kata "hypertext" dalam HTML berarti "teks" yang dapat berfungsi sebagai link dan halaman biasa. Gambar juga dapat berfungsi sebagai link dan halaman lain.

Markup adalah istilah HTML tambahan yang berasal dari kata bahasa Inggris "tanda atau penanda". Untuk mengatur dan menyusun halaman web, HTML

menggunakan banyak karakter khusus. Tag informasi yang disampaikan ke browser adalah bagian dari kode HTML ini (Pratama, 2016).

Kata ketiga, berarti "bahasa". Menurut Pratama (2016), HTML adalah bukan bahasa pemrograman karena tidak memiliki struktur dasar seperti *logic* dan *object*.

Dari ini menunjukkan bahwa HTML adalah kumpulan simbol atau tag yang ditulis dalam suatu file dan dimaksudkan untuk menampilkan informasi yang ditulis dalam bentuk simbol atau pada halaman penjelajahan web (Duckett, 2011).

Tag ini memberikan informasi kepada browser yang menampilkan seluruh halaman web kepada pengguna. Tag HTML selalu diawali dengan tag pembuka dan penutup (Duckett, 2011).

Halaman web dikelilingi oleh tag pembuka dan penutup `</html>` untuk mulai menulis perintah HTML. dokumen HTML selalu diakhiri dengan ekstensi `*.htm` atau `*.html`.

2.10.2 PHP (Hypertext Preprocessor)

Menurut Internet Programming oleh Muhamad Hendra Sunarya dan Muhamad Bahid (2020, 34), PHP atau PHP merupakan singkatan rekursif: Hypertext Preprocessor merupakan bahasa pemrograman yang termasuk dalam kategori pemrograman sisi server (Muhardin, 2003). Pemrograman server-side adalah jenis bahasa pemrograman di mana server mengeksekusi skrip/program. Kemudian hasil pengolahan skrip/program dikirimkan ke klien sebagai output.

Selain Server-side programming, PHP juga memiliki pemrograman sisi klien. Jenis pemrograman ini merupakan kebalikan dari pemrograman sisi server. Untuk pemrograman sisi klien, skrip/program diproses di dalam klien.

2.10.3 CSS (Cascading Style Sheet)

Menurut Muhamad Hendra Sunarya dan Muhamad Bahid Internet Programming (2020, 10), CSS Cascading Style Sheet merupakan versi ketiga yang mengatur dan mengontrol tampilan suatu blog/halaman web. CSS awalnya dibuat, dikembangkan dan distandarisasi oleh W3C pada tahun 1996 (Robbins, 2012).

CSS Cascading Style Sheet sebagai membangun tampilan yang lebih menarik dan rapih. Ada empat untuk memasang kode CSS ke dalam HTML / Halaman Web yaitu:

1. Inline style sheet, mengacu pada penggunaan gaya CSS yang disisipkan langsung ke dalam elemen HTML individual menggunakan atribut style. Ini berarti gaya CSS ditulis langsung di dalam tag HTML yang relevan, dan gaya ini hanya berlaku untuk elemen tersebut secara spesifik.
2. Internal style sheet, adalah metode untuk menentukan gaya CSS di dalam dokumen HTML yang sama, di dalam bagian `<style>` yang terletak di antara tag `<head>` dan `</head>`. Ini memungkinkan Anda untuk mendefinisikan gaya-gaya yang akan diterapkan pada elemen-elemen di seluruh halaman web tersebut.
3. Melakukan referensi ke dokumen dengan ekstensi `*.css`.
4. Impor file CSS mengimpor CSS ke file lain dengan tag `import`.

Artinya “CSS adalah singkatan dari Cascading Style Sheets. Secara umum sering disebut juga bahasa pemrograman stylesheet yang digunakan untuk mendefinisikan tampilan dan format dari sebuah dokumen yang ditulis dalam bahasa markup seperti HTML atau XML. CSS memungkinkan pengembang web

untuk mengontrol presentasi visual elemen-elemen halaman web, termasuk warna, ukuran teks, layout, dan berbagai aspek tampilan lainnya".

2.10.4 Java

Menurut (Ali Subhan, 2017), "Java adalah bahasa pemrograman yang dapat dipakai untuk menyusun program pada berbagai sistem operasi. Java adalah bahasa pemrograman yang umum digunakan (Firly 2018). Pengembang dapat menggunakan IDE yang mendukung Java, seperti Netbeans, Eclipse, dan TexPad, tetapi Java tidak memiliki IDE khusus. Simbol, tag, kata kunci, dan tipe data primitif adalah elemen dasar pemrograman Java.

2.10.5 Javascript

Javascript adalah bahasa pemrograman yang sering digunakan untuk mengembangkan aplikasi web interaktif. Dieksekusi di sisi klien (client-side) dalam peramban web, JavaScript memungkinkan interaksi dengan elemen HTML, mengontrol perilaku halaman web, dan berkomunikasi dengan server untuk memperbarui konten tanpa memuat ulang halaman. Sebagai bahasa skrip Internet pertama, JavaScript menyediakan fungsionalitas tambahan pada HTML dengan memungkinkan eksekusi perintah di sisi pengguna (browser) alih-alih di sisi server web (Cohen dalam Wahyudi dkk, 2017) (Marisa 2017:42).

2.11 Perangkat Lunak (Software)

Software adalah program yang dijalankan oleh komputer agar dapat melakukan tugas sesuai keinginan pengguna (Kadir 2017:2). Contoh perangkat lunak termasuk aplikasi seperti Microsoft Word dan Microsoft Excel, serta sistem operasi seperti Windows, Mac OS, dan Linux. Menurut Sabargunani (2009:65)

Kadir, penting untuk mendefinisikan perangkat lunak atau software dalam dua cara utama lakukan sendiri atau sudah jadi.

Menurut Mulyani dalam (Fitriadi, 2018), Perangkat Lunak (Software) adalah suatu perangkat yang tidak mempunyai bentuk fisik tetapi dapat dioperasikan oleh seorang pengguna atau user melalui seperangkat program komputer yang digunakan untuk mengolah data”.

2.11.1 Visual Studio Code

Visual Studio Code menggunakan komponen editor yang digunakan dalam Visual Studio Team Service (sebelumnya Visual Studio Online) yang bernama “Monaco” (Lardinois, 2015). Sebagaimana dinyatakan oleh Rulli (2017), Visual Studio merupakan sebuah *software* yang lengkap untuk mengembangkan sebuah aplikasi.

2.11.2 XAMPP

XAMPP merupakan sebuah paket perangkat lunak yang digunakan untuk membuat lingkungan pengembangan web lokal. XAMPP mencakup server web Apache, database MySQL, serta bahasa pemrograman PHP dan Perl yang dibutuhkan untuk pengembangan aplikasi web (Riyanto dalam Isty & Afifah 2018) Kesuma & Kholifah, 2019).

Raharjo (2018:20) menambahkan XAMPP memberikan kemudahan untuk menginstal paket perangkat lunak yang diperlukan untuk proses pengembangan aplikasi web tanpa harus melakukannya secara terpisah.

2.11.3 Framework

Framework merupakan kumpulan kode berupa perpustakaan dan *tools* yang

disatukan menjadi satu kerangka kerja untuk membantu proses pengembangan aplikasi (Budi Raharjo 2015:2).

“Framework adalah suatu kerangka kerja atau sekumpulan file yang berisi, dimana file tersebut berisi perintah-perintah dan fungsi-fungsi dasar kode program untuk melakukan tugas-tugas tertentu”, kata Purbadian (2016:14).

2.11.4 Bootstrap

Alat dalam, Bootstrap adalah struktur atau alat yang berfungsi untuk membantu mempercepat pengembangan web. Ini terdiri dari *styling* untuk membuat layout, grid, tipografi, tabel, formulir, navigasi, dan fitur lainnya (Wahyudi dkk 2017) (Nugroho & Setiyawati, 2019).

Framework ini memungkinkan pengembang web untuk membangun situs web responsif dengan cepat, karena telah menyediakan kumpulan aturan CSS dan JavaScript yang dapat digunakan untuk mempercantik tampilan dan meningkatkan fungsionalitas situs web.

2.11.5 JQuery

Menurut Hidayatullah, Kawistara (2017:395) beliau mengatakan dalam bukunya Web Programming bahwa “JQuery merupakan sebuah pustaka JavaScript yang dapat memperbaiki UI dan stabilitas website Anda, mempercepat waktu dan efisiensi dalam pembuatan website, dan hanya memanggil fungsi tanpa membuatnya dari awal”.

2.11.6 Basis Data

Menurut Rahimi Fitri (2020, 1) basis data merupakan kumpulan informasi yang dapat disimpan dan diakses oleh sistem komputer. *Software* yang dapat berfungsi untuk mengelola basis data disebut *Database Management System* atau

DBMS. DBMS merupakan *software* yang dibuat untuk mengelola database dan memproses data dengan lebih mudah. RDBMS, atau Relational DBMS merupakan salah satu jenis DBMS yang memungkinkan untuk memiliki relasi antar tabel. Contoh RDBMS termasuk MySQL dan PostgreSQL.

2.11.7 MySQL

Pemrograman Basis Data Menggunakan MySQL, kata Rahimi Fitri (2020, 2) Teknologi, termasuk perangkat lunak, berkembang pesat seiring berjalannya waktu. Softwaranya termasuk MySQL, yang selalu diupdate oleh produsennya. MySQL dimulai sebagai bagian dari proyek UNIREG.

MySQL adalah sistem database open source yang dilisensikan kedalam perangkat lunak bebas (perangkat lunak gratis) dan perangkat lunak berbagi (perangkat lunak dengan penggunaan terbatas). Oleh karena itu, MySQL adalah server basis data gratis yang dilisensikan oleh GNU General Public License (GPL), sehingga dapat digunakan untuk tujuan pribadi atau bisnis tanpa perlu membayar lisensi.. MySQL merupakan DBMS multithread dan multipengguna yang mendukung database bahasa SQL sebagai sistem manajemen data interaktif.

2.11.8 Perbedaan MySQL dan SQL

SQL merupakan bahasa pemrograman yang dipakai untuk mengatur database, sedangkan MySQL merupakan merek perangkat lunak sistem manajemen basis data (DBMS) yang dimaksudkan untuk mengatur database menggunakan bahasa SQL.

2.12 Web Browser

Browser memungkinkan pengguna untuk menampilkan dan menyelusuri situs web, serta untuk mengakses berbagai jenis konten seperti teks, gambar, video, dan

aplikasi web lainnya. Browser juga mendukung interaksi dengan berbagai layanan internet seperti email, media sosial, dan berbagai aplikasi web lainnya (Supardi, dkk 2017:168).

2.13 UML (Unifed Modeling Language)

Menurut Yuni Sugiarti (2018, 100) Peranti lunak saat ini tidak dapat lagi dibuat secara asal-asalan karena semakin luas dan luas. Oleh karena itu, hal-hal seperti skalabilitas, keamanan, dan eksekusi yang kuat harus dipertimbangkan saat membuat peranti lunak. Selain itu, arsitekturnya harus didefinisikan dengan jelas sehingga orang lain selain pengembang aslinya dapat dengan mudah menemukan dan memperbaiki. Keuntungan dari perancangan arsitektur yang cermat adalah modul atau komponen dapat digunakan kembali untuk program lain yang memiliki fungsi yang sama (Sugiarti, 2013).

Modeling adalah proses merancang *software* sebelum pengkodeannya. Membuat rencana pembangunan rumah adalah analogi yang baik untuk model perangkat lunak. Sangat penting untuk memodelkan sistem yang lebih kompleks karena kita belum memahami sistem sepenuhnya.

Dengan menggunakan model ini, pengembangan perangkat lunak diharapkan dapat sepenuhnya memenuhi kebutuhan pengguna, seperti skalabilitas, keandalan, dan keamanan, antara lain.

2.13.1 Use Case Diagram

Yuni Sugiarti (2018, 109) Saat membuat suatu sistem, langkah pertama merupakan menentukan kebutuhan fungsional dan kebutuhan non fungsional. Persyaratan fungsional merupakan kebutuhan sehari-hari pengguna sistem dan pemangku kepentingan, dimana pengguna dan pemangku kepentingan

menggunakan kebutuhan tersebut. Kebutuhan non-fungsional adalah kebutuhan yang mempertimbangkan hal-hal berikut yaitu kinerja, kegunaan sistem, keandalan sistem, keamanan sistem, keekonomian, legalitas dan pengoperasian.

Salah satu cara untuk menggambarkan kebutuhan fungsional adalah dengan menggunakan diagram tugas, juga dikenal sebagai diagram tugas. Diagram tugas memvisualisasikan hubungan antara satu atau lebih aktor dan sistem yang dibuat. Diagram *use case* memungkinkan dengan cepat mengetahui fitur yang tersedia dalam sistem dan aktor yang memiliki hak untuk menggunakannya. Marini, M. (2019), menjelaskan bahwa use case diagram menggambarkan kebutuhan sistem yang diperlukan dan menggambarkan secara jelas pengguna yang akan menggunakan sistem dan bagaimana pengguna dapat berinteraksi dengan sistem. Komponen Use Case Diagram ditampilkan pada tabel 2.1 dibawah ini :

Tabel 2.1 Komponen Use Case Diagram Sumber: (Rosa & Shalahuddin, 2018)

No.	Simbol	Deskripsi
1.	Use Case	Use case menggambarkan bagaimana pengguna atau aktor akan berinteraksi dengan sistem untuk mencapai suatu tujuan atau fungsi tertentu. Ini termasuk langkah-langkah spesifik yang dilakukan oleh pengguna dan respons yang diharapkan dari sistem.

2.	<p>Aktor/<i>Actor</i></p> 	<p>Dalam setiap use case, terdapat setidaknya satu aktor yang terlibat. Aktor ini bisa berupa pengguna langsung dari sistem (seperti pengguna akhir) atau sistem lain atau komponen eksternal yang berinteraksi dengan sistem yang sedang dianalisis.</p>
3.	<p>Asosiasi / <i>Assosiation</i></p>	<p>Abstraksi dari penghubung antara aktor dengan use case</p>
4.	<p><i>Include</i> <code><<include>></code></p>	<p>use case sepenuhnya merupakan fungsi dari use case lain.</p>
5.	<p>Ekstensi / <i>Extend</i></p>	<p>hubungan antara use case tambahan dan sebuah use case yang independen tanpa use case tambahan itu.</p>
6.	<p>Generalisasi</p>	<p>Hubungan generalisasi-spesialisasi merupakan tambahan fungsional dari use case lainnya jika suatu kondisi terpenuhifungsional dari use case lainnya jika suatu kondisi penuh</p>

2.13.2 Activity Diagram

Yuni Sugiarti (2018, 133) Diagram aktivitas atau activity diagram digambarkan dalam diagram aktivitas, juga dikenal sebagai diagram aktivitas. Diagram ini menunjukkan fungsi yang dijalankan sistem. Selain membantu

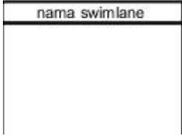
perilaku paralel, diagram aktivitas juga digunakan untuk menjelaskan hal-hal berikut :

1. merencanakan proses bisnis, yaitu setiap proses kegiatan menunjukkan proses sistem bisnis yang dibuat.
2. urutan atau pengelompokan antarmuka pengguna atau sistem di mana setiap kegiatan dianggap memiliki antarmuka pengguna; dan.
3. Perencanaan pengujian di mana setiap kegiatan dianggap memerlukan sebuah pengujian yang harus menjelaskan kasus ujiannya.

Dalam (Marini, M. 2019), menjelaskan bahwa diagram fungsi mengvisualisasikan proses atau fungsi suatu sistem dalam suatu perangkat lunak. Pada tahap perancangan diagram aktivitas, dijelaskan setiap fungsi desain use case. Komponen Activity diagram akan di jelaskan pada tabel dibawah ini:

Tabel 2.2 Kompone Anctivity Diagram Sumber: (Rosa & Shaludin, 2013)

Simbol	Definisi
Status awal 	Posisi awal pengoperasian sistem, diagram kerja mempunyai keadaan awal.
Aktivitas 	Operasi sistem biasanya dimulai dengan kata kerja.
Percabangan / <i>decision</i> 	Cabang yang terhubung ke lebih dari satu garis tindakan.
Penggabungan / <i>join</i> 	Bisnis yang menggabungkan lebih dari satu bisnis menjadi satu.

<p>Status akhir</p> 	<p>Dalam diagram aktivitas, keadaan akhir adalah keadaan akhir yang diselesaikan sistem.</p>
<p><i>Swimlane</i></p> 	<p>memisahkan organisasi bisnis yang bertanggung jawab atas operasi.</p>

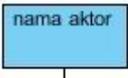
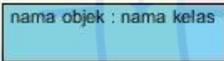
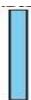
2.13.3 Squance Diagram

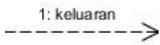
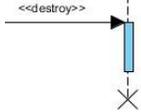
Yuni Sugiarti (2018, 130) Sequence diagram mengvisualisasikan aksi objek dalam use case dengan menampilkan masa hidup objek dan pesan yang dikirim dan diterima antar objek. Untuk membuat diagram sequence perlu mengetahui objek yang berhubungan dengan use case dan juga metode/fungsi kelas yang dipakai pada objek tersebut.

Semakin banyak pendefinisian Usecase yang memiliki proses sendiri atau yang penting, semakin banyak diagram sekuens yang harus dibuat. Akibatnya, semakin banyak pendefinisian Usecase, semakin banyak diagram sekuens yang harus dibuat.

Menurut (Irmayani & Susyatih, 2017), diagram ini menjelaskan proses dari sistem merespons tindakan pengguna. Diagram ini dibuat memiliki hubungan langsung dengan fungsi utama sistem. Komponen Sequence Diagram ditampilkan pada tabel 2.3 dibawah ini :

Tabel 2.3 Komponen Sequence Diagram Sumber : (Rosa & Shalahddin, 2014:165)

Simbol	Keterangan
<p>Aktor</p>  <p>nama aktor</p> <p>Atau</p> 	<p>Pengguna, proses, atau sistem lain yang berinteraksi dengan sistem.</p>
<p>Garis hidup / <i>Lifeline</i></p> 	<p>Menggambarkan kehidupan suatu objek</p>
<p>Objek</p> 	<p>Mendefinisikan objek yang berinteraksi pesan.</p>
<p>Waktu aktif</p> 	<p>mendefinisikan objek dalam keadaan aktif dan berinteraksi dengannya.</p>
<p>Pesan tipe <i>create</i></p>  <p><<<i>create</i>>></p>	<p>Saat menyatakan suatu objek, arah panah mengarah pada objek yang dibuat.</p>
<p>Pesan tipe <i>call</i></p> 	<p>Menunjukkan bahwa suatu objek yang menggabungkan fungsi yang sudah ada pada objek lain atau dirinya sendiri.</p>

<p>Pesan tipe <i>send</i></p> 	<p>Menunjukkan bahwa suatu objek mengirimkan pesan ke objek lainnya.</p>
<p>Pesan tipe <i>return</i></p> 	<p>Menunjukkan suatu objek yang mengembalikan hasil proses ke objek lainnya.</p>
<p>Pesan tipe <i>destroy</i></p> 	<p>Untuk menunjukkan bahwa objek mengakhiri hidup objek yang lain.</p>

