# BAB II
# LANDASAN TEORI

Bab ini menjelaskan tentang dasar teori yang digunakan dalam menyusun tugas akhir ini, yaitu tentang pengertian bank, pengertian kredit, unsur-unsur kredit, pengertian data mining.

## 2.1 Pengertian Bank

Definisi bank menurut Undang-undang Nomor 10 Tahun 1998 sebagaimana telah diubah dengan Undang-undang Nomor 7 Tahun 1992, tentang Perbankan, pada Pasal 1 disebutkan bahwa : "Bank adalah badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak"[1].

**Fungsi Bank, yaitu :**

1. Untuk menciptakan uang. Uang yang diciptakan bank umum adalah uang giral, yaitu alat pembayaran lewat mekanisme pemindahbukuan (kliring). Kemampuan bank umum menciptakan uang giral menyebabkan posisi dan fungsinya dalam pelaksanaan kebijakan moneter. Bank sentral dapat mengurangi atau menambah jumlah uang yang beredar dengan cara mempengaruhi kemampuan bank umum menciptakan uang giral[2].
2. Mendukung Kelancaran Mekanisme Pembayaran Fungsi lain dari bank umum yang juga sangat penting adalah mendukung kelancaran mekanisme pembayaran. Hal ini dimungkinkan karena salah satu jasa yang ditawarkan bank umum adalah jasa-jasa yang berkaitan dengan mekanisme pembayaran. Beberapa jasa yang amat dikenal adalah kliring, transfer uang, penerimaan setoran-setoran, pemberian fasilitas pembayaran dengan tunai, kredit, fasilitas-fasilitas pembayaran yang mudah dan nyaman, seperti kartu plastik dan sistem pembayaran elektronik[2].
3. Penghimpunan Dana Simpanan Masyarakat Dana yang paling banyak dihimpun oleh bank umum adalah dana simpanan. Di Indonesia dana

simpanan terdiri atas giro, deposito berjangka, sertifikat deposito, tabungan dan atau bentuk lainnya yang dapat dipersamakan dengan itu. Kemampuan bank umum menghimpun dana jauh lebih besar dibandingkan dengan lembaga-lembaga keuangan lainnya. Dana-dana simpanan yang berhasil dihimpun akan disalurkan kepada pihak-pihak yang membutuhkan, utamanya melalui penyaluran kredit[2].

4. Mendukung Kelancaran Transaksi Internasional Bank umum juga sangat dibutuhkan untuk memudahkan dan atau memperlancar transaksi internasional, baik transaksi barang/jasa maupun transaksi modal. Kesulitan-kesulitan transaksi antara dua pihak yang berbeda negara selalu muncul karena perbedaan geografis, jarak, budaya dan sistem moneter masing-masing negara. Kehadiran memudahkan penyelesaian transaksi-transaksi tersebut. Dengan adanya bank umum yang beroperasi dalam skala internasional akan bank umum, kepentingan pihak-pihak yang melakukan transaksi transaksi internasional dapat ditangani dengan lebih mudah, cepat, dan murah[2].

## 2.2 Pengertian Kredit

Istilah kredit bermakna percaya, karena itu dasar dari kata kredit adalah kepercayaan bahwa seseorang atau penerima kredit akan menyanggupi segala sesuatu yang telah diperjanjikan terlebih dahulu pada masa yang akan datang[2].

Kredit dalam pengertian ekonomi merupakan suatu penundaan pembayaran, yaitu uang atau barang (prestasi) yang diterima sekarang akan dikembalikan pada masa yang akan datang berikut tambahan suatu kontra prestasi[2].

Secara ringkas kredit bermakna “suatu pemberian prestasi dari suatu pihak kepada pihak lain dan prestasi itu akan dikembalikan lagi pada suatu masa tertentu yang akan datang disertai dengan suatu kontra prestasi berupa bunga”. berdasarkan Undang-Undang Pokok Perbankan No.14 tahun 1967, Kredit diuraikan sebagai “penyediaan uang atau tagihan-tagihan yang dapat disamakan, dengan itu berlandaskan persetujuan pinjam-meminjam antara bank dengan pihak

lain. Berdasarkan definisi kredit, maka pihak peminjam berkewajiban melunasi hutang sesudah masa waktu tertentu dengan jumlah bunga yang telah ditetapkan" (kasmar, 2006) [2].

### 2.2.1 Unsur-unsur Kredit

Pada dasarnya pemberian kredit didasarkan atas kepercayaan, yang berarti bahwa pemberian kepercayaan oleh bank sebagai pemberi kredit, dimana prestasi yang diberikan benar-benar sudah diyakini akan dapat dibayar kembali oleh penerima kredit sesuai dengan syarat-syarat yang telah disetujui bersama. Berdasarkan hal-hal tersebut dapat ditarik kesimpulan bahwa unsur-unsur yang terdapat dalam pemberian kredit adalah[2]:

1. Kepercayaan, yaitu keyakinan si pemberi kredit (bank) bahwa prestasi (uang) yang diberikan akan benar-benarkembali dari penerima kredit pada suatu masa yang akan datang[2]..
2. Waktu, yaitu jangka waktu antara saat pemberian prestasi dengan saat pengembaliannya. Dalam unsur waktu ini terkandung pengertian tentang nilai agio yaitu nlai uang sekarang lebih berharga daripada nilai uang dimasa yang akan datang, sehingga dalam hal ini perlu adanya kontra prestasi yang harus berupa uang[2].
3. Resiko, yaitu resiko yang dapat timbul pada saat peberian kredit. Untuk menghindari resiko, maka sebelum kredit diberikan harus dilakukan penilaian secara cermat dan dilindungi dengan agunan atau jaminan kredit sebagai benteng terakhir dalam pengaman kredit. Penilaian berdasarkan atas banafiditas calon penerima kredit sehingga dapat dapat ditentukan sampai sejauh mana calon penerima kredit dapat dipercaya oleh bank[2].

Prestasi, dalam hubungannya dengan pemberian kredit yang dimaksud prestasi adalah uang[2].

### 2.2.2 Tujuan Kredit

Tujuan kredit dilihat dari segi tujuan pemakaiannya adalah:

1. Kredit Produktif

   Merupakan kredit yang digunakan untuk meningkatkan usaha atau produksi atau investasi. Kredit ini diberikan untuk menghasilkan barang dan jasa. Artinya kredit ini digunakan untuk diusahakan sehingga menghasilkan suatu baik berupa barang maupun jasa[2].
2. Kredit Konsumtif

   Merupakan kredit yang digunakan untuk dikonsumsi atau dipakai secara pribadi. Dalam kredit ini tidak ada pertambahan barang dan jasa yang dihasilkan, karena memang untuk digunakan atau dipakai oleh seseorang atau badan usaha[2].

### 2.2.3 Prosedur Kredit

Prosedur pengajuan kredit di sebuah PT Bank Negeri yang ada di Kota Jailolo Kabupaten Halmahera Barat Provinsi Maluku Utara, bagi calon nasabah yang ingin mengajukan permohonan kredit, maka nasabah tersebut harus melalui prosedur sebagai berikut[2]:

1. Nasabah mengambil formulir pengajuan kredit di bank yang bersangkutan[4].
2. Nasabah mengisi formulir tersebut, formulir pengajuan kredit terdiri dari :
   a. Data pemohon kredit di antaranya, nama lengkap dan gelar, nama/alias, tempat/tanggal lahir, jenis kelamin, nama gadis ibu kandung, nomor KTP/SIM/Pasport pemohon, status sipil, pendidikan, hubungan dengan bank, alamat pemohon kepemilikan rumah, lama menetap, nama dan alamat yang dihubungi dalam keadaan darurat, nomor telepon/*handphone*, nomor fax, NPWP[2].
   b. Data pemohon Kredit di antaranya, kebutuhan kredit, tujuan penggunaan, kemampuan angsuran per-bulan, dan jangka waktu kredit[2].
   c. Data pekerjaan di antaranya, nama Instansi/Perusahaan, alamat Instansi/Perusahaan, nomor telepon, nomor fax, bidang usaha, jabatan,

mulai bekerja, usia pension, nama atasan langsung, penghasilan perbulan[2].

d. Data Keluarga di antaranya, nama istri/suami, tempat/tanggal lahir, alamat, nomor telepon/nomor fax, pekerjaan, pengahasilan, perbulan, keluarga yang ditanggung[2].
e. Data Jaminan/Objek yang dibiayai di antaranya, rumah (alamat, luas (tanah/bangunan), nilai (Jt), An), kendaraan (jenis kendaraan, nomor Polisi, nomor BPKB, tahun, An, nilai (Jt)), lainnya (jenis jaminan, lokasi, An, nilai (Jt)) [2].
f. Kondisi Keuangan Keluarga
    a) Kekayaan Pribadi di antaranya, deposito, tabungan giro, saham, nilai tunai asuransi jiwa, tanah dan bangunan, kendaraan dan lain-lain[4].
    b) Hutang diantaranya, jenis pinjaman, sisa hutang, masa berlaku, kreditur, agunan[2].
    c) Sumber Dana di antaranya (Pendapatan keluarga setiap bulan) diantaranya, sumber penghasilan (gaji yang diterima, usaha tambahan sendiri, penghasilan istri/suami) [2].
    d) Penggunaan (Pengeluaran Biaya) diantaranya, sewa pemeliharaan rumah, biaya hidup, perawatan kendaraan, premi asuransi, tunjangan kepada pihak ketiga, angsuran/tabungan, pengeluaran lain-lain[2].
    e) Sisa pendapatan bersih[2].

3. Nasabah melengkapi persyaratan yang telah ditentukan, adapun persyaratan tersebut antara lain[2] :
    a. Foto copy kartu tanda penduduk suami/istri
    b. Foto copy surat nikah/cerai
    c. Foto copy kartu keluarga
    d. Suarat kuasa memotong gaji/pensiunan yang disetujui oleh bandahara dan pimpinan instansi/lembaga/perusahaan
    e. Surat persetujuan dari suami/istri
    f. Surat rekomendasi dari Pimpinan/Kepala Dinas
    g. Surat pernyataan pembayaran angsuran di atas materai

h. Foto copy kartu pegawai dan Taspen
i. Daftar gaji yang ditandatangani oleh Bendahara Gaji dan Atasan

4. Nasabah yang telah mengisi formulir dan melengkapi persyaratan kemudian menyerahkan formulir dan persyaratan tersebut ke pihak Bank bagian kredit[2].

## 2.3 Data Mining

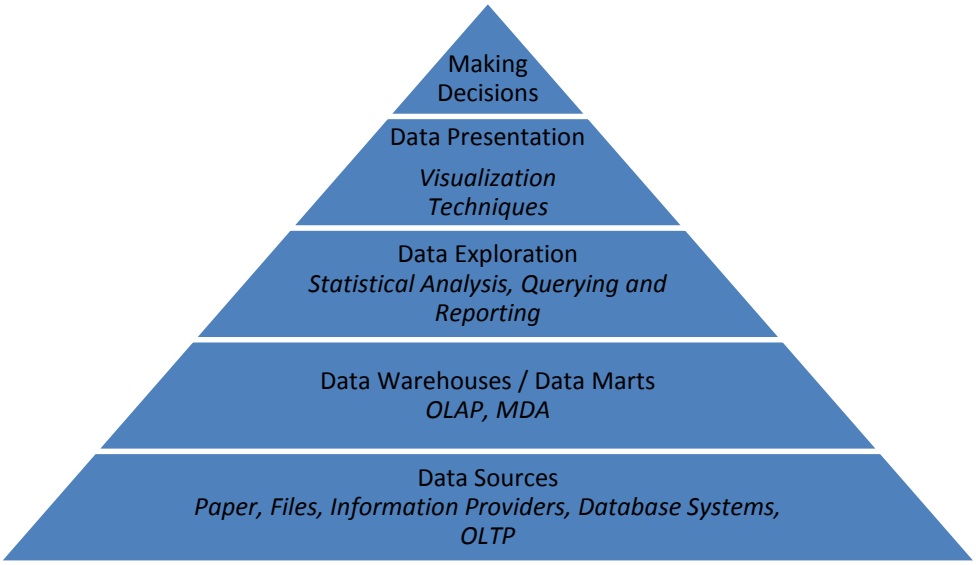
*Data mining* adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *Knowledge Discovery in Databases* (KDD) adalah penerapan metode saintifik pada *data mining.* Dalam konteks ini *data mining* merupakan satu langkah dari proses KDD[3].

*Data Mining* merupakan proses iteratif dan interaktif untuk menemukan pola atau model baru yang sahih (sempurna), bermanfaat dan dapat dimengerti dalam suatu *database* yang sangat besar (*massive database*) [3].

- Sahih : Dapat digeneralisasi untuk masa yang akan datang.
- Baru : Apa yang sedang tidak diketahui.
- Bermanfaat : Dapat digunakan untuk melakukan suatu tindakan.
- Iteratif : Memerlukan sejumlah proses yang diulang.
- Interaktif : Memerlukan interaksi manusia dalam prosesnya.

*Data mining* berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya[3].

Contoh yang bukan *Data mining* adalah mencari nomor telpon dalam direktori telepon atau *query* suatu *Web search engine* untuk informasi mengenai "Amazon". Sedangkan yang termasuk dalam *data mining* adalah mencari nama-nama khusus yang lazim di lokasi tertentu seperti di US (O'Brien, O'Reilly… di daerah Boston) dan mengelompokkan dokumen-dokumen yang sama yang diperoleh dari *search engine* menurut konteksnya (misal. Amazon rainforest, Amazon.com) [3].




*Gambar 0-1 Data Mining & Teknologi Database Lainnya.*

Dari gambar di atas terlihat bahwa teknologi *data warehouse* digunakan untuk melakukan OLAP, sedangkan *data mining* digunakan untuk melakukan *information discovery* yang informasinya lebih ditujukan untuk seorang *Data Analyst* dan *Business Analyst* (dengan ditambah visualisasi tentunya). Dalam prakteknya, *data mining* juga mengambil data dari *data warehouse*. Hanya saja aplikasi dari *data mining* lebih khusus dan lebih spesifik dibandingkan OLAP mengingat *database* bukan satu-satunya bidang ilmu yang mempengaruhi *data mining*, banyak lagi bidang ilmu yang turut memperkaya *data mining* seperti: *information science* (ilmu informasi), *high performance computing*, visualisasi, *machine learning*, statistic, *neural networks* (jaringan syaraf tiruan), pemodelan matematika, *information retrieval* dan *information extraction* serta pengenalan

pola. Bahkan pengolahan citra (*image Processing*) juga digunakan dalam rangka melakukan *data mining* terhadap data *image/spatial*[3].

Alasan mengapa melakukan *data mining* dari sudut pandang komersial karena (1) meledaknya volume data yang dihimpun dan disimpan dalam data *warehouse* seperti data *web, e-commerce,* penjualan di *department store,* transaksi bank/*credit card,* (2) proses komputasi yang dapat diupayakan, (3) kuatnya tekanan kompetitif untuk dapat menyediakan yang lebih baik, layanan-layanan *custom*-isasi dan informasi sedang menjadi produk yang berarti[3].

Alasan dari sudut pandang keilmuan adalah (1) kecepatan data yang dihimpun dan disimpan (GB/hour) seperti pada *remote sensors pada satellite, telescopes* untuk memindai langit dan simulasi santifik yang membangkitkan data dalam ukuran terabytes, (2) teknik-teknik trasional tidak fisibel untuk mengolah data mentah, (3) *data mining* untuk reduksi data untuk klasifikasi dan segmentasi data serta membantu ilmuwan dalam melakukan formasi hipotesis[3].

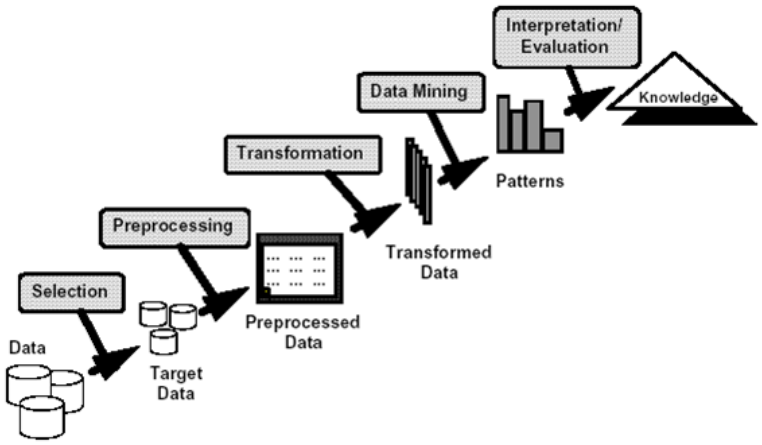
Alasan mula *data mining* mengambil ide dari *machine learning*/AI, pengenalan pola, statistic, dan *database systems.* Teknik-teknik tradisional mungkin tidak sesuai karena (1) membludaknya data (*Enormity of data*), (2) dimensi data yang tinggi (*High dimensionality of data),* (3) heterogenitas dan sifat data yang tersebar (*Heterogeneous, distributed nature of data*) [3].

### 2.3.1 Operasi Data Mining

Operasi *data mining* menurut sifatnya dibedakan menjadi dua, yaitu bersifat (1) prediksi (*prediction driven*) untuk menjawab pertanyaan apa dan suatu yang bersifat remang-remang atau transparan. Operasi prediksi digunakan untuk validasi hipotesis, *querying* dan pelaporan (missal: *spreadsheet* dan *pivot table*), analisis multidimensi (*dimensional summar*); OLAP (*Online Analyst Processing*) serta analisis statistic. (2) penemuan (*discovery driven*) bersifat transparan dan untuk menjawab pertanyaan "mengapa?" Operasi penemuan digunakan untuk analisis data eksplorasi, pemodelan prediksif, segmentasi *database, analisis* keterkaitan (*link analysis*) dan deteksi devisi[3].

Tahapan proses dalam penggunaan *data mining* yang perupakan proses *Knowledge Discovery* in *databases* (KDD) seperti pada Gambar II-2 dapat diuraikan sebagai berikut[3]:

- Memahami *domain* aplikasi untuk mengetahui dan menggali pengetahuan awal serta apa sasaran pengguna.
- Membuat target data-set yang meliputi pemulihan data dan fokus pada sub-set data.
- Pembersihan dan transformasi data meliputi eliminasi derau, *outliers, missing value* serta pemilihan fitur dan reduksi dimensi.
- Penggunaan algoritma *data mining* yang terdiri dari asosiasi, sekuensial, klasifikasi klasterisasi dan lain-lain.
- Interpretasi, evaluasi dan visualisasi pola untuk melihat apakah ada sesuatu yang baru dan menarik dan dilakukan iterasi jika diperlukan.




*Gambar 0-2 Proses KDD*

### 2.3.2 Permasalahan Dalam Data Mining

Sistem *data mining* berdasar pada basis data yang menyediakan data mentah dan ini memunculkan permasalahan dalam basis data yang cenderung dinamis, tidak lengkap, ber-noise dan besar. Permasalahan lain muncul sebagai akibat dari kecukupan dan relevansi dari informasi yang disimpan[3].

Basis data seringkali didesain untuk tujuan yang berbeda dari *data mining* dan kadangkala properti atau atribut yang akan menyederhanakan pekerjaan pembelajaran tidak tersedia atau tidak dapat dimintai dari dunia nyata. Data yang tidak meyakinkan menyebabkan permasalahan karena jika ada atribut-atribut esensial bagi pengetahuan tentang domain aplikasi tidak ada dalam data tidak memungkinkan untuk menemukan pengetahuan yang tepat mengenai domain yang diberikan. Sebagai contoh, kita tidak dapat mendiaknosa malaria dari basis data pasien jika basis data tersebut tidak mengandung jumlah sel darah merah pasien[3].

Basis data biasanya dicemari oleh *error* sedemikian hingga dapat diasumsikan bahwa data secara keseluruhan benar. Atribut-atribut yang ada pada subjek atau pertimbangan ukuran dapat memunculkan kesalahan (*error*) sedemikian hingga beberapa contoh mungkin menjadi mis-klasifikasi. *Error* dalam salah satu nilai atribut atau informasi kelas dikenal sebagai *noise.* Secara nyata ada kemungkinan kita perlu sekali untuk menghilangkan *noise* dari informasi klasifikasi saat hal ini mempengaruhi akurasi aturan yang dibangkitkan secara keseluruhan[3].

Data yang dapat dibenahi dengan sistem penemuan dalam berbagai cara, seperti[3]:

- Secara sederhana dengan mengabaikan nilai-nilai yang hilang.
- Menghilangkan *record* yang berhubungan.
- Menebak nilai yang menghilang dari nilai-nilai yang diketahui.
- Memperlakukan data yang hilang sebagai sebuah nilai khusus yang dimasukkan sebagai tambahan dalam domain atribut.
- Atau menghitung rata-rata nilai yang hilang menggunakan Bayesian.

Data yang mengandung *noise* dalam pengertian menjadi tidak teliti merupakan karakteristik dari semua koleksi data dan secara khusus cocok untuk sebuah distribusi statistik biasa seperti *Gaussian* saat nilai-nilai yang salah merupakan kesalahan masukan data[3].

Ketidakjelasan (*uncertainty)* menunjuk kepada kepemilikan *error* dan tingkat *noise* dalam data. Presisi data merupakan salah satu pertimbangan penting dalam sistem penemuan[3].

Basis data cenderung menjadi besar dan dinamis dalam hal isinya yang selalu berubah saat informasi ditambahkan, dimodifikasi atau dihapus. Permasalahan dalam hal ini dari sudut pandang *data mining* adalah bagaimana menjamin bahwa aturan-aturan tersebut *up-to-date* dan konsisten dengan informasi paling terkini. Juga sistem pembelajaran mempunyai *time-sensitive* saat beberapa nilai data berubah terhadap waktu dan sistem penemuan dipengaruhi oleh ketepatan dari data tersebut[3].

### 2.3.3 Teknik Data Mining

Ada beberapa teknik yang dimiliki *data mining* berdasarkan Kusrini dkk, 2006, yaitu :

*1. Deskripsi*

Para peneliti/analis biasanya mencoba menemukan cara untuk mendeskripsikan pola dan trend yang tersembunyi dalam data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

*2. Estimasi*

Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih ke arah numerik daripada kategori. Misalnya, akan dilakukan estimasi tekanan *systolic* dari pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah.

*3. Prediksi*

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi dimasa depan). Misalnya, ingin diketahui prediksi harga beras tiga bulan yang akan datang.

*4. Klasifikasi*

Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, kita akan mengklasifikasikan pendapatan dalam tiga kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

*5. Clustering*

*Custering* lebih kearah pengelompokan record, pengamatan, atau kasus dalam kelas yang memiliki kemiripan. Sebuah *cluster* adalah kumpulan record yang memiliki kemiripan satu dengan yang lain dan memiliki ketidak miripan dengan record-record dalam *cluster* yang lain, misalnya untuk tujuan audit akuntansi akan dilakukan segmentasi perilaku finansial dalam kategori dan mencurigakan.

## 2.4 Konsep Klasifikasi

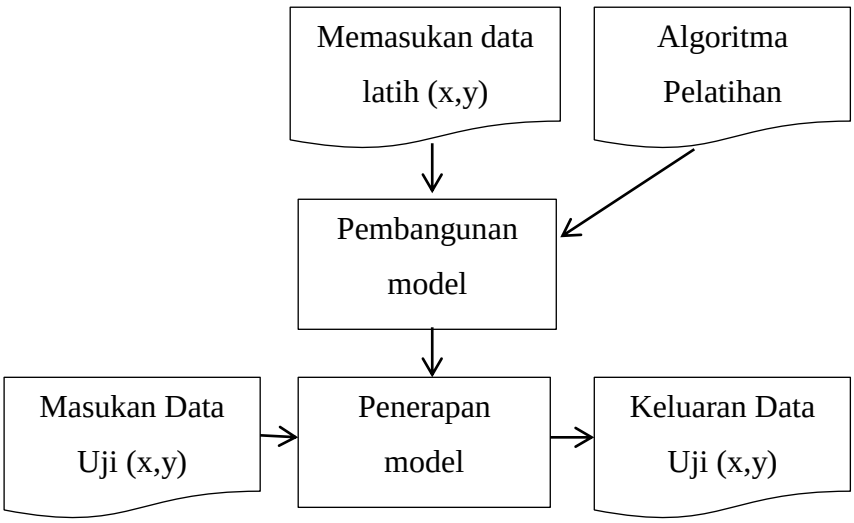
Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu (1) pembangunan model sebagai prototype untuk disimpan sebagai memori dan (2) penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya[4].

Contoh aplikasi yang sering ditemui adalah pengklasifikasian jenis hewan, yang mempunyai sejumlah atribut. Dengan atribut tersebut, jika ada hewan baru, kelas hewannya bisa langsung di ketahui. Contoh lain adalah bagaimana melakukan diagnosis penyakit kulit kanker melanoma (Amaliyah *et al*, 2022), yaitu dengan melakukan pembangunan model berdasarkan data latih yang ada, kemudian menggunakan model tersebut untuk mengidentifikasi penyakit pasien baru sehingga diketahui apakah pasien tersebut menderita kanker atau tidak[4].

### 2.4.1 Model

Klasifikasi dapat didefinisikan sebagai pekerjaan yang melakukan pelatihan/pembelajaran terhadap fungsi target *f* yang memetakan setiap set attribute (fitur) *x* ke satu dari sejumlah label kelas *y* yang tersedia. Pekerjaan pelatihan tersebut akan menghasilkan suatu model yang kemudian disimpan sebagai memori[4].

Model dalam klasifikasi mempunyai arti yang sama dengan kotak hitam, di mana ada suatu model yang menerima masukan, kemudian mampu melakukan memikiran terhadap masukan tersebut, dan memberikan jawaban sebagai keluaran dari hasil pemikirannya. Kerangka kerja (*framework*) klasifikasi ditunjukan pada Gambar II-3. Pada gambar tersebut disediakan sejumlah data latih (x,y) untuk digunakan sebagai data pembangun model. Model tersebut kemudian dipakai untuk memprediksi kelas dari data uji (x,?) sehingga diketahui kelas *y* yang sesungguhnya[4].




*Gambar 0-3 Proses Pekerjaan Klasifikasi*

Model yang sudah dibangun pada saat pelatihan kemudian dapat digunakan untuk memprediksi label kelas data baru yang belum diketahui. Dalam pembangunan model selama proses pelatihan tersebut diperlukan suatu algoritma untuk membangunnya, yang disebut algoritma pelatihan (*learning algorithm*).

Ada banyak algoritma pelatihan yang sudah dikembangkan oleh banyak peneliti, seperti *K-Nearest Neighbor, Artificial Neural Network, Support Vector Machine,* dan sebagainya. Setiap algoritma mempunyai kelebihan dan kekurangan, tetapi semua algoritma berprinsip sama, yaitu melakukan suatu pelatihan sehingga di akhir pelatihan, model dapat memetakan (memprediksi) setiap vektor masukan ke label kelas keluaran dengan benar[4].

Kerangka kerja seperti yang ditunjukan pada gambar diatas pakai nomor meliputi dua langkah proses, yaitu induksi dan deduksi. Induksi merupakan langkah untuk membangun model klasifikasi dari data latih yang diberikan, disebut juga proses pelatihan, sedangkan deduksi merupakan langkah untuk menerapkan model tersebut pada data uji sehingga kelas yang sesungguhnya dari data uji dapat diketahui, disebut juga proses prediksi[4].

Berdasarkan cara pelatihan, algoritma-algoritma dapat dibagi menjadi dua macam, yaitu *eager learner* dan *lazy learner.* Algoritma-algoritma yang termasuk dalam kategori *eager learner* didesain untuk melakukan pembacaan/pelatihan/pembelajaran pada data latih agar dapat memetakan dengan benar setiap vektor masukan ke label kelas keluarannya sehingga di akhir proses pelatihan, model sudah dapat memetakan semua vektor data uji ke label kelas keluarannya dengan benar. Selanjutnya, setelah proses pelatihan tersebut selesai, model (biasanya berupa bobot atau sejumlah nilai kuantitas tertentu) disimpan sebagai memori, sedangkan semua data latihnya dibuang. Proses prediksi dilakukan dengan model yang tersimpan, tidak melibatkan data latih sama sekali. Cara ini mengakibatkan proses prediksi bejalan dengan cepat, tetapi harus dibayar dengan proses pelatihan yang lama. Algoritma-algoritma klasifikasi yang masuk kategori ini, di antaranya, adalah *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM), *Decision Tree,* Bayesian, dan sebagainya[4].

Sementara algoritma-algoritma yang masuk dalam kategori *lazy learner* hanya sedikit melakukan pelatihan (atau tidak sama sekali), hanya menyimpan sebagian atau seluruh data latih, kemudian melakukannya dalam proses prediksi. Hal ini mengakibatkan proses prediksi menjadi lama karena model harus membaca kembali semua data latihnya agar dapat memberikan keluaran label

kelas dengan benar pada data uji yang yang diberikan. Kelebihan algoritma seperti ini adalah proses pelatihan yang berjalan dengan cepat. Algoritma-algoritma klasifikasi yang masuk kategori ini, di antaranya, adalah *K-Nearest Neighbor* (K-NN), *Fauzy K-Nearest Neighbor* (FKNN), Regresi linear, dan sebagainya[4].

### 2.4.2 Pengukuran Kinerja Klasifikasi

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua set data dengan benar, tetapi tidak dapat dimungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya. Umumnya, pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*) [4].

Matriks komfusi merupakan tabel pencatat hasil kerja klasifikasi. Tabel II.1 merupakan contoh matriks konfusi yang melakukan klasifikasi masalah biner (dua kelas), hanya ada dua kelas, yaitu kelas 0 dan 1. Setiap sel $f_{ij}$ dalam matriks menyatakan jumlah rekord/data dari kelas *i* yang hasil prediksinya masuk ke kelas *j*. Misalnya, sel $f_{11}$ adalah jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1, dan $f_{10}$ adalah data dalam kelas 1 yang dipetakan secara salah ke kelas 0[4].

Tabel 0-1 Matriks konfusi untuk klasifikasi dua kelas

| $f_{ij}$ | | Kelas hasil prediksi (*j*) | |
|---|---|---|---|
| | | Kelas = 1 | Kelas = 0 |
| Kelas asli (*i*) | Kelas = 1 | $f_{11}$ | $f_{10}$ |
| | Kelas = 0 | $F_{01}$ | $f_{00}$ |

Berdasarkan matriks konfusi, kita dapat mengetahui jumlah data dari masing-masing kelas yang diprediksi secara benar, yaitu ($f_{11} + f_{00}$), dan data yang diklasifikasikan secara salah, yaitu ($f_{01} + f_{10}$). Kuantitas matriks konfusi dapat diringkas menjadi dua nilai, yaitu akurasi dan laju eror. Dengan mengetahui jumlah data yang diklasifikasikan secara benar, kita dapat mengetahui akurasi hasil prediksi, dan dengan mengetahui jumlah data yang diklasifikasikan secara

salah, kita dapat mengetahui laju eror dari prediksi yang dilakukan. Dua kuantitas ini digunakan sebagai matrik kinerja klasifikasi[4].

Untuk menghitung akurasi digunakan formula

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} = \frac{f11 + f00}{f11+f10+f01+f00}$$

Untuk menghitung laju eror (kesalahan prediksi) digunakan formula

$$\text{Laju eror} = \frac{\text{Jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} = \frac{f10 + f01}{f11+f10+f01+f00}$$

Semua algoritma klasifikasi berusaha membentuk model yang mempunyai akurasi tinggi (laju eror yang rendah). Umumnya, model yang dibangun dapat mempredikasi dengan benar pada semua data yang menjadi data latihnya, tetapi ketika model berhadapan dengan data uji, barulah kinerja model dari sebuah algoritma klasifikasi ditentukan[4].

## 2.5 C5.0

Algoritma C5.0 merupakan penyempurnaan dari algoritma terdahulu yang dibentuk oleh Ross Quinlan pada tahun 1987, yaitu algoritma ID3 dan C4.5. Perbedaan utama C5.0 dari ID3 adalah[5]:

- C5.0 dapat menangani atribut kontinyu dan diskrit. Akan tetapi pada penelitian ini yang dipakai hanya atribut diskrit karena jika menggunakan atribut kontinyu pohon yang dihasilkan akan panjang dan tidak efisien[5].
- Hasil pohon keputusan C5.0 dapat dipangkas atau terdapat pruning (pemangkasan). Atribut dengan nilai information gain tertinggi akan terpilih sebagai *parent* bagi *node* selanjutnya. Algoritma ini membentuk pohon keputusan dengan cara pembagian dan menguasai sampel secara rekursif dari atas ke bawah. Algoritma ini dimulai dengan semua data yang dijadikan akar dari pohon keputusan sedangkan atribut yang dipilih akan menjadi pembagi bagi sampel tersebut[5].

$$I\,(s1\,,\,s2\,,\,\ldots\,,\,sm) = -\sum_{i=1}^{m} \text{pi log2 (pi)}$$

S adalah sebuah himpunan yang terdiri dari s data sampel. Diketahui atribut class adalah m dimana mendefinisikan kelas-kelas di dalamnya, Ci (for i= 1,…,m), si adalah jumlah sampel pada S dalam class Ci. untuk mengklasifikasikan sampel yang digunakan maka diperlukan informasi dengan menggunakan aturan 3. Dimana pi adalah proporsi kelas dalam output seperti pada kelas Ci dan diestimasikan dengan si /s. Atribut A memiliki nilai tertentu {a1, a2,…, av}. Atribut A dapat[5].

digunakan pada partisi S ke dalam v subset, {S1, S2, …, Sv}, dimana Sj berisi sampel pada S yang bernilai aj pada A. Jika A dipilih sebagai atribut tes (sebagai contoh atribut terbaik untuk split), maka subset ini akan berhubungan pada cabang dari node himpunan S. Sij adalah jumlah sampel pada class Ci dalam sebuah subset Sj. Untuk mendapatkan informasi nilai subset dari atribut A tersebut maka digunakan formula[5],

$$E\ (A) = \sum_{j=1}^{y} \frac{s1j+\ldots+smj}{s}\ I\ (S1j, \ldots, Smj)$$

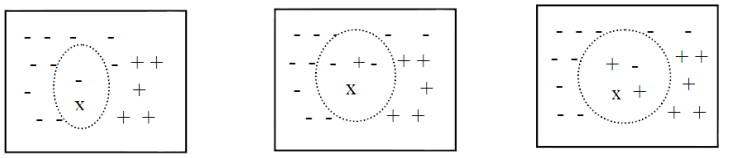
$\frac{s1j+\ldots+smj}{s}$ adalah jumlah subset j yang dibagi dengan jumlah sampel pada S, maka untuk mendapatkan nilai gain, selanjutnya digunakan formula[5],

$$Gain\ (A) = I\ (s_1\ , s_2\ , \ldots\ , s_m) - E(A)$$

## 2.6 Nearest-Neighbor Classifer

K-Nearest Neighbor Algorithm Seperti halnya decision tree, K-Nearest Neighbor sangat sering digunakan dalam klasifikasi dengan tujuan dari algoritme ini adalah untuk mengklasifikasi objek baru berdasarkan atribut dan training *samples* (Larose, 2002 ). Algoritme k-nearest neighbor (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Teknik ini sangat sederhana dan mudah diimplementasikan. Mirip dengan teknik klastering, pengelompokkan suatu data baru berdasarkan jarak data baru itu ke beberapa data/tetangga (*neighbor*) terdekat. Dalam hal ini jumlah data/tetangga terdekat ditentukan oleh *user* yang dinyatakan dengan *k*. Misalkan ditentukan *k*=5, maka

setiap *data testing* dihitung jaraknya terhadap *data training* dan dipilih 5 *data training* yang jaraknya paling dekat ke *data testing*. Lalu periksa *output* atau labelnya masing-masing, kemudian tentukan *output* mana yang frekuensinya paling banyak. Lalu masukkan suatu *data testing* ke kelompok dengan *output* paling banyak. Misalkan dalam kasus klasifikasi dengan 3 kelas, lima data tadi terbagi atas tiga data dengan *output* kelas 1, satu data dengan *output* kelas 2 dan satu data dengan *output* kelas 3, maka dapat disimpulkan bahwa *output* dengan label kelas 1 adalah yang paling banyak. Maka data baru tadi dapat dikelompokkan ke dalam kelas 1. Prosedur ini dilakukan untuk semua *data testing* (Santosa, 2007). Gambar 4 berikut ini adalah bentuk representasi K-NN dengan 1, 2 dan 3 tetangga data terhadap data baru *x* (Pramudiono, 2003) [5].




(a)1-*nearest neighbor* (b)2-*nearest neighbor* (c)3-*nearest neighbor*

*Gambar 0-4 Ilustrasi 1-, 2-, 3-nearest neighbor terhadap data baru (x)*

Untuk mendefinisikan jarak antara dua titik yaitu titik pada *data training* (x) dan titik pada *data testing* (y) maka digunakan rumus Euclidean,

$$d_{(x,y)}\sqrt{\sum_{i=1}^{n}(xi - yi)^2}$$

Dengan *d* adalah jarak antara titik pada data *training x* dan titik data *testing y* yang akan diklasifikasi, dimana $x=x_1,x_2,\ldots,x_i$ dan $y=y_1,y_2,\ldots,y_i$ dan *I* merepresentasikan nilai atribut serta *n* merupakan dimensi atribut (Han & Kamber, 2001). Sebagai ilustrasi, pada Tabel II.2 berikut ini disajikan contoh penerapan rumus Euclidean, pada empat data klasifikasi kualitas baik dan tidak baik sebuah kertas tisu yang dinilai berdasarkan daya tahan kertas tersebut dan fungsinya. Sebanyak tiga data yang sudah terklasifikasi yaitu data no 1,2, dan 3 masing-masing data dihitung jaraknya ke data

no 4 untuk mendapatkan kelas yang sesuai bagi data no 4 maka *k*=1 (Teknomo, 2006) [5].

Tabel 0-2 Klasifikasi kualitas baik atau tidak baik sebuah kertas tisu

| No | Fungsi | Daya Tahan | Klasifikasi |
|---|---|---|---|
| 1 | 7 | 7 | Tidak baik |
| 2 | 7 | 4 | Tidak baik |
| 3 | 3 | 4 | Baik |
| 4 | 1 | 4 | ? |

Berikut ini disajikan pula perhitungan yang dilakukan terhadap tiga data yang sudah terklasifikasi dengan data yang belum terklasifikasi pada Tabel di atas[5].

Jarak data no satu ke data no empat:

$d_{1,4}\sqrt{(7-1)^2+(7-4)^2} = \sqrt{6^2-6^2} = \sqrt{45} = 6.07$

Jarak data no dua ke data no empat:

$d_{2,4}\sqrt{(7-1)^2+(4-4)^2} = \sqrt{6^2-0^2} = \sqrt{36} = 6$

Jarak data no tiga ke data no empat:

$d_{3,4}\sqrt{(3-1)^2+(4-4)^2} = \sqrt{2^2-0^2} = \sqrt{4} = 2$

Dari hasil perhitungan di atas diperoleh jarak antara data no tiga dan data no empat adalah jarak yang terdekat maka kelas data no empat adalah baik. Teknik ini akan diujicobakan terhadap *dataset* akademik yang belum terklasifikasi atau data yang belum dikenal, untuk menemukan kelas yang sesuai dengan berdasarkan pada data tetangga terdekatnya yang sudah terklasifikasi. Tingkat ketepatan klasifikasi terhadap data dari kedua algoritma yang digunakan menjadi titik fokus analisa dalam penelitian[5].

## 2.7 UML

UML adalah (Unified Modeling Language) bahasa untuk menspesifikasi, memvisualisasi, membangun dan mendokumentasikan *artifacts* (bagian dari informasi yang digunakan untuk dihasilkan oleh proses pembuatan perangkat lunak, *artifact* tersebut dapat berupa model, deskripsi atau perangkat lunak)dari sistem perangkat lunak, seperti pada pemodelan bisnis dan sistem non perangkat lunak lainnya. Selain itu UML adalah bahasa pemodelan yang menggunakan konsep orientasi *object*. UML dibuat oleh Grady Booch, James Rumbaugh, dan Ivar Jacobson di bawah bendera *Rational Software Corps*. UML menyediakan notasi-notasi yang membantu memodelkan sistem dari berbagai prespetktif. UML tidak hanya digunakan dalam pemodelan perangkat lunak, namun hampir dalam semua bidang yang membutuhkan pemodelan [6].




*Gambar 0-5 Logo UML*

### 2.7.1 Use Case Diagram

Use Case merupakan sebuah teknik yang digunakan dalam pengembangan sebuah software atau sistem informasi untuk menangkap kebutuhan fungsional dari sistem yang bersangkutan, Use Case menjelaskan interaksi yang terjadi antara 'aktor'—inisiator dari interaksi sistem itu sendiri dengan sistem yang ada, sebuah Use Case direpresentasikan dengan urutan langkah yang sederhana[7].

Perilaku sistem adalah bagaimana sistem beraksi dan bereaksi. Perilaku ini merupakan aktifitas sistem yang bisa dilihat dari luar dan bisa diuji.Perilaku

sistem ini dicapture di dalam USE CASE. USE CASE sendiri mendeskripsikan sistem, lingkungan sistem, serta hubungan antara sistem dengan lingkungannya[7].

### 2.7.2 Sequence Diagram

Sequence diagram adalah diagram yang menggambarkan kolaborasi dinamis antara sejumlah object. Kegunaannya untuk menunjukkan rangkaian pesan yang dikirim antara object juga interaksi antara object. Sesuatu yang terjadi pada titik tertentu dalam eksekusi sistem.Activity Diagram[8].

### 2.7.3 Class Diagram

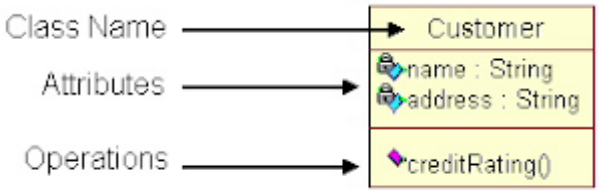
Diagram kelas atau class diagram menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem. Kelas memiliki apa yang disebut atribut dan metode atau operasi [9].

atribut merupakan variabel-variabel yang dimiliki oleh suatu kelas

operasi atau metode adalah fungsi-fungsi yang dimiliki oleh suatu kelas

Diagram kelas mendeskripsikan jenis-jenis objek dalam sistem dan berbagai hubungan statis yang terdapat di antara mereka. Diagram kelas juga menunjukkan properti dan operasi sebuah kelas dan batasan-batasan yang terdapat dalam hubungan-hubungan objek tersebut. [9].

Diagram kelas menggambarkan struktur dan deskripsi *class*, *package* dan objek beserta hubungan satu sama lain seperti containment, pewarisan, asosiasi, dan lain-lain. [9].



Kelas memiliki tiga area pokok :

1. Nama
2. Atribut
3. Operasi

Contoh **kelas** : Manusia

**Atribut** : nama, usia, tanggal lahir

**Method/Operasi** : berjalan, makan, minum

### 2.7.4 Entity Relationship Diagram (ERD)

*Entity Relationship Diagram* (ERD) adalah suatu model untuk menjelaskan mengenai hubungan antar data dalam basis data berdasarkan objek-objek dasar data yang memiliki hubungan antar relasi. Entity Relationship Diagram oyi untuk memodelkan struktur data serta hubungan antar data, untuk dapat menggambarkannya digunakan beberapa notasi serta simbol[10].

Menurut Brady serta Loonam (2010), Entity Relationship diagram adalah suatu teknik yang digunakan untuk dapat memodelkan kebutuhan data dari sebuah organisasi, biasanya oleh System Analys didalam tahap analisis persyaratan proyek pengembangan system[10].

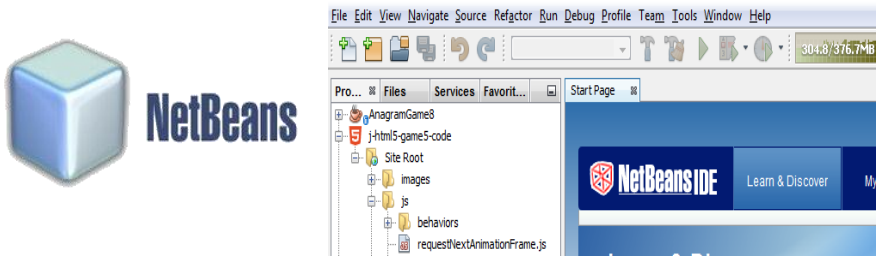
## 2.8 Software Pendukung

*Software* pendukung merupakan alat yang digunakan untuk membangun sistem atau aplikasi yang digunakan. *Software* pendukung yang digunakan yaitu NetBeans IDE 8.2, Rstudio, XAMPP Control Panel v3.2.2 dan Microsoft Excel

### 2.8.1 NetBeans IDE

Netbeans merupakan sebuah aplikasi Integrated Development Environment (IDE) yang berbasiskan Java dari Sun Microsystems yang berjalan di atas swing. Swing merupakan sebuah teknologi Java untuk pengembangan aplikasi desktop yang dapat berjalan pada berbagai macam platform seperti

windows, linux, Mac OS X dan Solaris. Sebuah IDE merupakan lingkup pemrograman yang di integrasikan ke dalam suatu aplikasi perangkat lunak yang menyediakan Graphic User Interface (GUI), suatu kode editor atau text, suatu compiler dan suatu debugger[11].

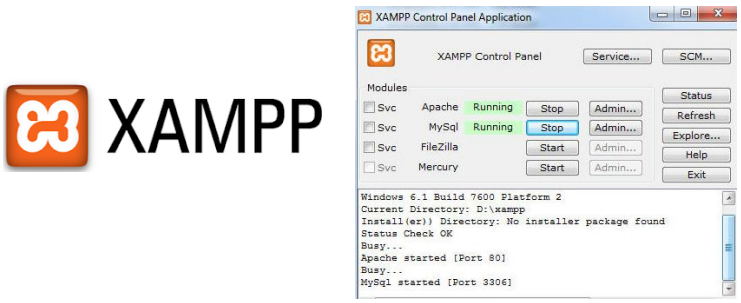



*Gambar 0-6 NetBeans IDE*

### 2.8.2 XAMPP

XAMPP ialah perangkat lunak bebas yang mendukung banyak sistem operasi, merupakan campuran dari beberapa program. Yang mempunyai fungsi sebagai server yang berdiri sendiri (*localhost*), yang terdiri dari program MySQL *database*, Apache HTTP *Server*, dan penerjemah ditulis dalam bahasa pemrograman PHP dan Perl[12].

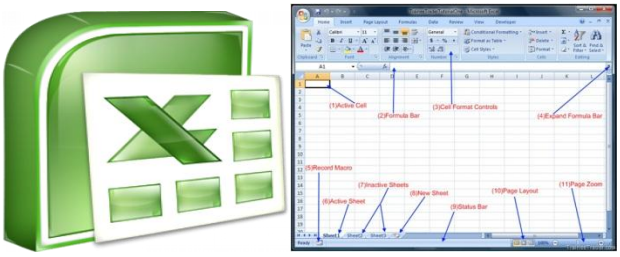
Nama XAMPP merupakan singkatan dari X (empat sistem operasi), Apache, MySQL, PHP dan Perl. Program ini tersedia di bawah GNU *General Public License* dan bebas, adalah mudah untuk menggunakan *web server* yang dapat melayani tampilan halaman web yang dinamis[12].




*Gambar 0-7 XAMPP*

### 2.8.3 Microsoft Exel

*Microsoft Excel* adalah sebuah program dari *Microsoft Office* yang mempunyai fungsi pengolah data berupa perhitungan atau grafik[13]..



*Gambar 0-8 Microsoft Excel*

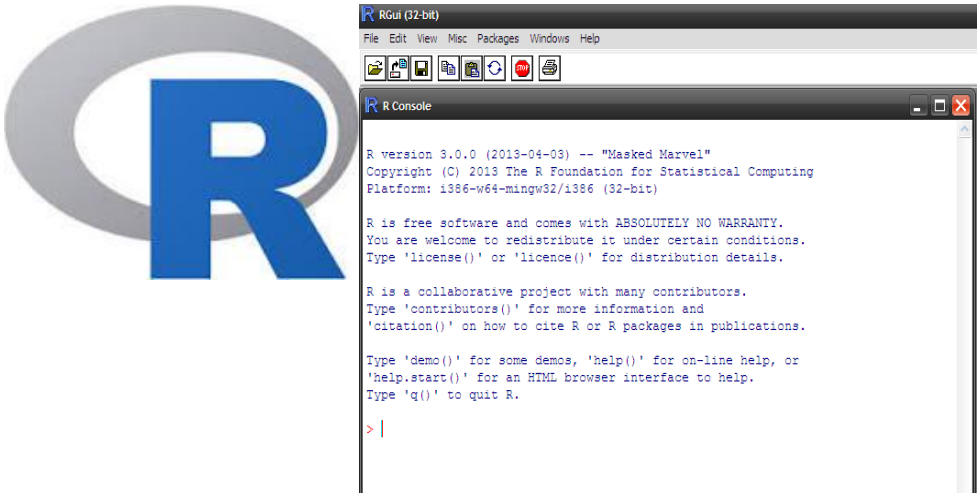
Program ini memungkinkan penggunanya untuk membuat lembar kerja *spreadsheet* dengan berbagai fitur kalkulasi akurat dan mudah dioperasikan[13].

Hingga saat ini, *Microsoft Excel* menjadi salah satu program *spreadsheet* terpopuler dan dipergunakan oleh berbagai kalangan, baik di platform Windows maupun Mac OS[13].

### 2.8.4 R

R (juga dikenal sebagai GNU S) adalah bahasa pemrograman dan perangkat lunak untuk analisis statistika dan grafik. R dibuat oleh Ross Ihaka dan Robert Gentleman di Universitas Auckland, Selandia Baru, dan kini dikembangkan oleh R Development Core Team, di mana Chambers merupakan anggotanya. R dinamakan sebagian setelah nama dua pembuatnya (Robert Gentleman dan Ross Ihaka), dan sebagian sebagian dari permainan nama dari S[14].

Bahasa R kini menjadi standar *de facto* di antara statistikawan untuk pengembangan perangkat lunak statistika, serta digunakan secara luas untuk pengembangan perangkat lunak statistika dan analisis data[14].




*Gambar 0-9 Bahasa Pemrograman R*

## 2.9 Pengujian Alpha

pengujian *Alpha* adalah salah satu strategi pengujian perangkat lunak yang paling umum digunakan dalam pengembangan perangkat lunak, hal ini khusus digunakan oleh organisasi pengembangan produk dengan tujuan agar system yang dikembangkan terhindar dari cacat atau kegagalan penggunaan[15].

Pengujian alpha berlangsung di situs pengembang oleh tim internal, sebelum rilis kepada pelanggan eksternal. Agar nantinya ketika pelanggan menggunakan system ini tidak kecewa karena masalah cacat atau kegagalan aplikasi. Pengujian ini dilakukan tanpa keterlibatan tim pengembangan[15].

## 2.10 Data Transformation: Normalization

Teknik yang digunakan untuk menormalisasikan data adalah dengan Min-Max, merupakan teknik yang membantu menormalkan data. Pada skala data antara 0 – 1. Berikut contoh normalisasi mengunakan Min-max[16].

| Nilai |
|---|
| 8 |
| 10 |
| 15 |
| 20 |

Min: Nilai minimal atribut yang diberikan. Disini Min adalah 8

Max: Nilai maksimal atribut yang diberikan. Disini Max adalah 20

$$v' = \frac{v - \min_A}{\max_A - \min_A}(new_\max_A - new_\min_A) + new_\min_A$$

V: V adalah nilai atribut masing-masing. Misalnya di sini V1 = 8, V2 = 10, V3 = 15 dan V4 = 20

newMax : 1

newMin : 0

Untuk Nilai Atribut = 8 :

$$MinMax = \frac{(V - Min\ marks)}{Max\ omarks - Min\ marks}(newMax - newMin)^{\square} + newMin$$

$$MinMax = \frac{(8 - 8)}{20 - 8} * (1 - 0)^{\square} + 0$$

$$MinMax = \frac{(0)}{12} * 1$$

$$MinMax = 0$$

Untuk Nilai Atribut = 10 :

$$MinMax = \frac{(10 - 8)}{20 - 8} * (1 - 0)^{\square} + 0$$

$$MinMax = \frac{(2)}{12} * 1$$

$$MinMax = 0.16$$

Untuk Nilai Atribut = 15 :

$$MinMax = \frac{(15 - 8)}{20 - 8} * (1 - 0)^{\square} + 0$$

$$MinMax = \frac{(3)}{12} * 1$$

$$MinMax = 0.25$$

Untuk Nilai Atribut = 20 :

$$MinMax = \frac{(20 - 8)}{20 - 8} * (1 - 0)^{\square} + 0$$

$$MinMax = \frac{(12)}{12} * 1$$

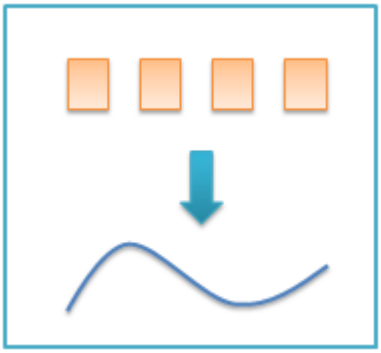
$$MinMax = 1$$

*Tabel 0-3* Contoh Min-Max Normalization

| **Nilai Atribut** | **Nilai Atribut Setelah Melakukan Min-Max Normalization** |
|---|---|
| 8 | 0 |
| 10 | 0.16 |
| 15 | 0.25 |
| 20 | 1 |

## 2.11 Teknik Encoding

Encoding atau continuization adalah transformasi variabel kategoris ke bilangan biner atau numerik. Contohnya adalah memperlakukan pria atau wanita untuk jenis kelamin sebagai 1 atau 0 [17].



*Gambar 0-10 Encoding*

Contoh encoding melalui data kategori.

*Tabel 0-4* Contoh encoding

| Categorical | Encoded | | |
|---|---|---|---|
| Trend | Trend_Up | Trend_Down | Trend_Flat |
| Up | 1 | 0 | 0 |
| Up | 1 | 0 | 0 |
| Down | 0 | 1 | 0 |
| Flat | 0 | 0 | 1 |
| Down | 0 | 1 | 0 |
| Up | 1 | 0 | 0 |
| Down | 0 | 1 | 0 |
| Flat | 0 | 0 | 1 |
| Flat | 0 | 0 | 1 |
| Flat | 0 | 0 | 1 |