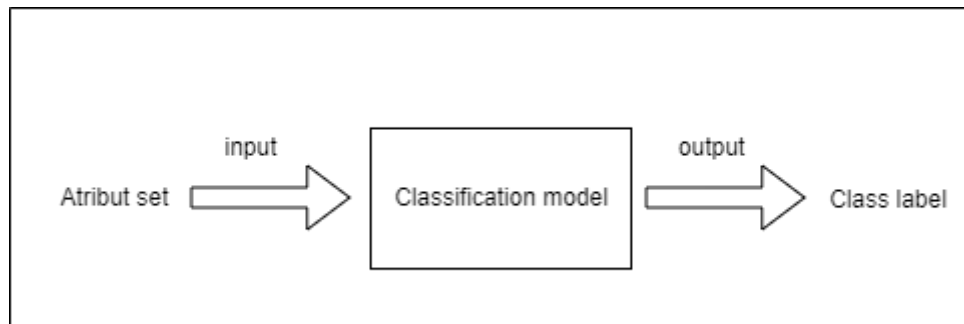


BAB II TINJAUAN PUSTAKA

2.1 Klasifikasi

Klasifikasi adalah proses dua langkah yaitu pembelajaran atau pelatihan (learning) dan klasifikasi (klasifikasi). Pembelajaran adalah langkah membuat model klasifikasi, dan klasifikasi adalah langkah menentukan kelas dari data yang disediakan menggunakan model yang dibuat (Hamid and Nurul Hidayat, 2019). Gambar 2.1 berikut merupakan proses tahapan dari klasifikasi:



Gambar 2.1 Tahapan Klasifikasi

Tahap pelatihan, suatu set data latih dengan kelas yang sudah diketahui dianalisis dan dibangun model dari setiap kelas dengan bantuan algoritme klasifikasi. Proses pelatihan pada klasifikasi disebut *supervised learning* karena setiap data latih sudah diketahui kelasnya masing-masing. Model yang di dapat merupakan aturan-aturan klasifikasi (*classification rules*). Aturan ini diuji dengan data uji untuk memperkirakan akurasi. Data uji bersifat independen dan tidak dipergunakan pada proses pelatihan. Jika akurasi baik, maka -aturan tersebut bisa digunakan untuk klasifikasi data baru yang kelasnya belum diketahui.

2.2 Penyakit Paru-Paru

Penyakit paru-paru adalah penyakit yang khusus menyerang organ paru-paru. Paru-paru merupakan salah satu organ pernapasan yang berfungsi untuk

melakukan repirasi, yaitu mengubah gas (CO_2) menjadi gas oksigen (O_2) dan air (H_2O). sebelum sampai ke paru-paru, udara yang dihirup manusia akan melewati hidung, pangkal laring, kemudian menuju ke kedua bronki utama (bronkus) dan akan disalurkan ke bronki yang paling kecil (bronkioli), selanjutnya udara dimasukkan ke dalam jutaan kantong udara (alveoli) yang berada dalam paru-paru. Paruparu memiliki lapisan pelindung (pleura) yang juga berfungsi untuk membantu kontraksi dalam rongga dada. Penyakit paru-paru adalah kondisi paru-paru dimana terjadi peradangan atau pengumpulan cairan (darah atau nana) atau masuknya bakteri, virus atau jamur ke dalam paru-pari yang kemudian menyebabkan paruparu tidak berfungsi dengan baik (Meiyanti and Komarudin, 2020).

2.3 Kecerdasan Buatan

Kecerdasan buatan berasal dari bahasa inggris "*Artificial Intellegence*" atau disingkat AI, yaitu *intelligence* adalah kata sifat yang berarti cerdas, sedangkan artificial artinya buatan. Kecerdasan buatan yang dimaksud di sini merujuk pada mesin yang mampu berfikir menimbang tindakan yang akan diambil, dan mampu mengambil keputusan seperti yang dilakukan oleh manusia.

Berikut adalah beberapa definisi kecerdasan buatan yang telah didefinisikan oleh beberapa ahli :

- a) Alan Turing, menetapkan definisi AI
Komputer tidak dapat dibedakan dengan manusia saat berbincang melalui terminal komputer, maka bias dikatakan komputer itu cerdas, mempunyai kecerdasan.
- b) Herbert Alexander Simon (june 15,1916-Februari 9, 2001)
Kecerdasan buatan merupakan kawasan penelitian, aplikasi dan intruksi yang terkait dengan pemograman komputer untuk melakukan sesuatu hal yang dalam pandangan manusia adalah cerdas.

c) Rich And Knight (1991)

“Kecerdasan buatan (AI) merupakan sebuah studi tentang bagaimana membuat computer melakukan hal-hal yang pada saat ini dapat dilakukan lebih baik oleh manusia”.

d) Encyclopedia Britannica

“Kecerdasan buatan (AI) merupakan cabang dari ilmu computer yang dalam mempresentasi pengetahuan lebih banyak menggunakan bentuk symbol-simbol daripada bilangan dan memproses informasi berdasarkan metode heuristik atau dengan berdasarkan jumlah aturan”.

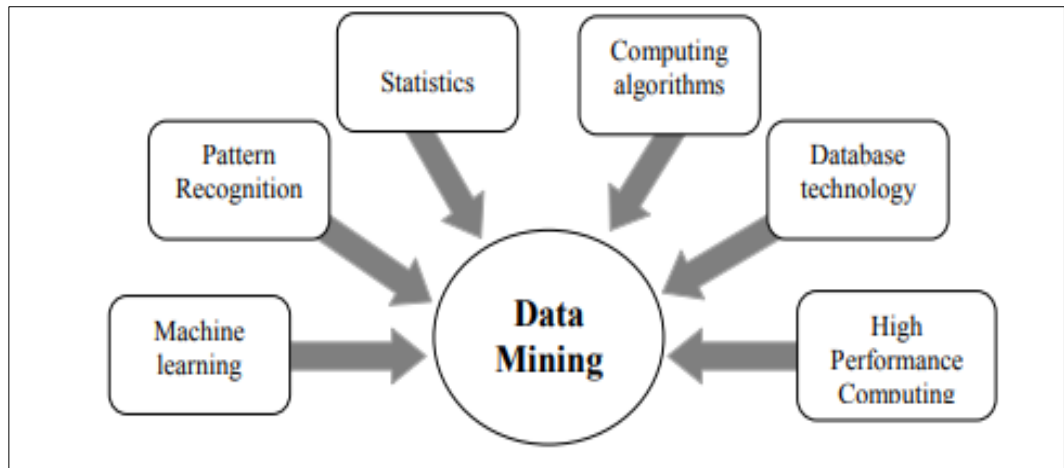
e) Menurut Winston dan Prendergast (1984)

- 1) Membuat mesin menjadi lebih pintar (tujuan utama)
- 2) Memahami apa itu kecerdasan (tujuan ilmiah)
- 3) Membuat mesin lebih bermanfaat (tujuan *entrepreneurial*)

2.4 Data Mining

Data mining dikenal sejak tahun 1990-an, ketika adanya suatu pekerjaan yang memanfaatkan data menjadi suatu hal yang lebih penting dalam berbagai bidang, seperti marketing dan bisnis, sains, dan teknik, serta seni dan hiburan. Sebagian ahli menyatakan bahwa *data mining* merupakan suatu langkah untuk menganalisis pengetahuan dalam basis data atau biasa disebut *Knowledge Discovery in Database* (KDD) . *Data mining* merupakan proses untuk menemukan pola data dan pengetahuan yang menarik dari kumpulan data yang sangat besar (Muslim *et al.*, 2019).

Data mining, secara sederhana merupakan suatu langkah ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan belum diketahui. *Data mining* mempunyai hubungan dengan berbagai bidang seperti statistic, machine learning, *computing algorithms*, *database technology*. Gambar 2.2 merupakan diagram hubungan *data mining* :



Gambar 2.2. Diagram Hubungan *Data Mining*

Secara sistematis, langkah utama untuk melakukan *data mining* terdiri dari tahap, yaitu sebagai berikut :

1) Ekspolasi Atau Pemrosesan Awal Data

Eksplorasi atau pemrosesan awal data terdiri dari pembersihan data, normalisasi data, transformasi data, penanganan missing value, reduksi dimensi, pemilihan subset fitur, dan sebagainya.

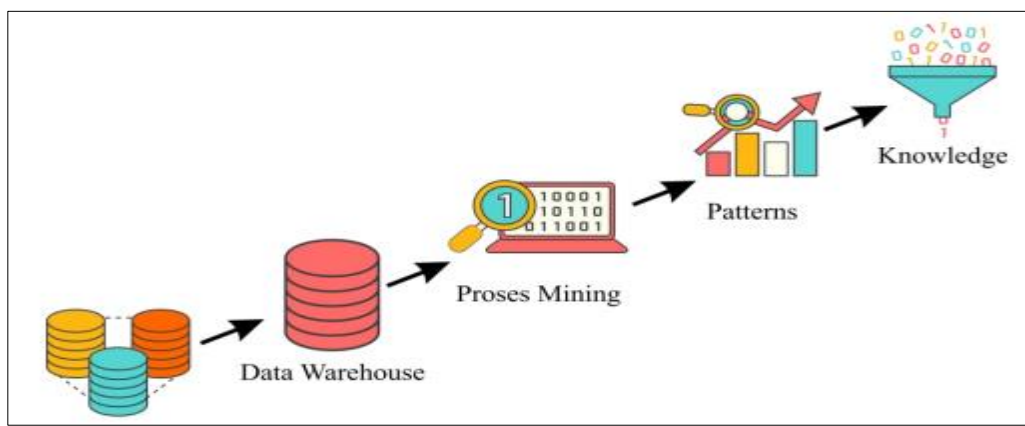
2) Membangun Model Dan Validasi

Membangun model dan validasi, merupakan melakukan analisis dari berbagai model dan memilih model sehingga menghasilkan kinerja yang terbaik. Pembangunan model dilakukan menggunakan metode-metode seperti klasifikasi, regresi, analisis cluster, dan asosiasi.

3) Penerapan

Penerapan dilakukan dengan menerapkan model yang dipilih pada data baru untuk menghasilkan kinerja yang baik pada masalah yang diinvestigasi.

Tahapan proses data mining ada beberapa yang sesuai dengan proses KDD (*Knowledge Discovery in Database*). Gambar 2.3 merupakan proses KDD (*Knowledge Discovery in Database*):



Gambar 2.3. Proses KDD (*Knowledge Discovery in Database*)

1. *Cleaning And Integration.*

a. *Data Cleaning* (Pembersih data)

Data cleaning (Pembersihan data) adalah proses yang dilakukan untuk menghilangkan noise pada data yang tidak konsisten atau bisa disebut tidak relevan. Data yang diperoleh dari database suatu perusahaan maupun hasil eksperimen yang sudah ada, tidak semuanya memiliki isian yang sempurna misalnya data yang hilang, data yang tidak valid, atau bisa juga hanya sekedar salah ketik. Data yang tidak relevan itu dapat ditangani dengan cara dibuang atau sering disebut dengan proses cleaning. Proses cleaning dapat berpengaruh terhadap performa dari teknik *data mining*.

b. *Data Integration* (Integrasi Data)

Integrasi data merupakan proses penggabungan data dari berbagai database sehingga menjadi satu database baru. Data yang diperlukan pada proses *data mining* tidak hanya berasal dari beberapa database.

2. *Selection and Transformation*

a. *Data Selection* (Seleksi Data)

Tidak semua data yang ada di database akan dipakai, karena hanya data yang sesuai saja yang akan dianalisis dan diambil dari database. Misalnya pada sebuah kasus market basket analysis yang akan meneliti faktor kecenderungan pelanggan, maka tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan.

b. *Data Transformation* (Transformasi Data)

Transformasi data merupakan proses perubahan data dan penggabungan data ke dalam format tertentu, *data mining* membutuhkan format data khusus sebelum diaplikasikan. Misalnya metode standar seperti analisis asosiasi dan clustering hanya bisa menerima inputan data yang bersifat kategorial. Karenanya data yang berupa angka numerik apabila mempunyai sifat kontinu perlu dibagi menjadi beberapa interval. Proses ini sering disebut dengan transformasi data.

3. *Poses Mining*

Proses mining dapat disebut juga sebagai proses penambangan data. Proses mining merupakan proses utama yang menggunakan metode untuk menemukan pengetahuan berharga yang tersembunyi dari data.

4. *Evaluation and Precentation*

a. Evaluasi Pola (*Pattren Evaluation*)

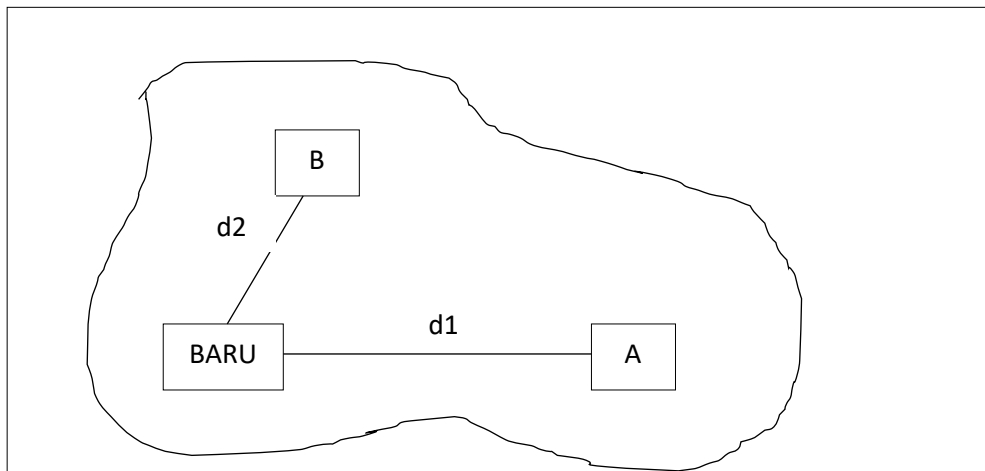
Evaluasi pola bertugas untuk mengidentifikasi pola-pola yang menarik ke dalam knowledge based yang ditemukan. Pada tahap ini dihasilkan polapola yang khas dari model klasifikasi yang dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai dengan hipotesa, terdapat beberapa alternative yang bias diambil seperti menjadikanya umpan baik untuk memprbaiki proses *data mining*, atau mencoba metode *data mining* lain yang lebih sesuai.

b. Presentasi Pengetahuan (*Knowledge Presentation*)

Knowledge presentation merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan atau informasi yang telah digali oleh pengguna. Tahap terakhir dari proses data mining adalah memformulasikan keputusan dari hasil analisis yang didapat.

2.5 **K-Nearest Neighbor (KNN)**

Metode *K-Nearest Neighbor* (KNN) adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pencocokan bobot dari sejumlah fitur yang ada. Misalkan diinginkan untuk mencari solusi terhadap seorang pasien baru dengan menggunakan solusi dari pasien terdahulu. Untuk mencari kasus mana yang akan digunakan, maka dihitung kedekatan kasus pasien baru dengan semua kasus pasien lama. Kasus pasien lama dengan kedekatan terbesar lah yang akan diambil solusinya untuk digunakan pada kasus pasien baru. Gambar 2.4 merupakan ilustrasi kedekatan kasus.



Gambar 2.4. Ilustrasi Kedekatan Kasus.

Seperti tampak pada gambar 2.4 ada dua pasien lama : A dan B. Ketika ada pasien Baru, maka solusi yang akan diambil adalah solusi dari pasien terdekat dari pasien Baru. Seandainya d1 adalah kedekatan antara pasien Baru dengan pasien A, sedangkan d2 adalah kedekatan antara pasien Baru dengan pasien B. Karena d2 lebih dekat dari d1, maka solusi dari pasien B-lah yang akan digunakan untuk memberikan solusi pasien Baru.

Berikut adalah persamaan *Euclidean Distance* dalam Algoritma K-NN :

$$k_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \dots\dots\dots (1)$$

Keterangan:

x_1 = Data Uji

x_2 = Data Training

i = Variabel Data

k = Jarak

p = Dimensi Data

2.6 Website

Website atau situs dapat diartikan sebagai kumpulan dari halaman-halaman yang digunakan untuk menampilkan informasi berupa teks, gambar, animasi, suara, dan atau gabungan dari semuanya baik yang bersifat statis maupun dinamis yang membentuk satu rangkaian bangunan yang saling terkait dan biasanya dibuat untuk perorangan, organisasi, dan perusahaan (Rahardja, Juardi and Agung, 2019).

2.7 Odoo Framework

Odoo *Framework* merupakan sebuah platform open source yang digunakan untuk keperluan bisnis. Aplikasi atau modul-modul yang terintegrasi dibangun di atas platform tersebut, meliputi semua bisnis dari CRM, Sales, stok, dan lain sebagainya. Odoo dibangun menggunakan bahasa pemrograman *Python*, XML, dan *JavaScript*. Odoo termasuk sebagai *software Enterprise Resource Planning (ERP)*. Odoo dulunya dikenal sebagai OpenERP. Odoo dibangun secara open source, sehingga Odoo mendukung pemanfaatan kembali library yang telah ada dan setiap orang dapat terlibat dalam pengembangannya. Platform Odoo terdiri dari tiga komponen utama, yaitu database *PostgreSQL* sebagai database bawaannya, application server Odoo, dan web server. Database *PostgreSQL* menampung semua data yang berhubungan dengan data dan konfigurasi Odoo. Selain sebagai aplikasi, Odoo juga dapat berfungsi sebagai *Framework* atau kerangka kerja bagi para Software Developer, (Daniel Reis, 2019).

2.8 Python

Python adalah bahasa pemrograman *multi-platform* yang bersifat *free* dan *open-source*, dan dapat digunakan untuk mengembangkan aplikasi-aplikasi desktop maupun web. Python memiliki pustaka standar (*Python Standard Library*) yang sangat lengkap sehingga dapat memenuhi berbagai macam permasalahan-permasalahan di dalam dunia pemrograman, sebagai alternatif

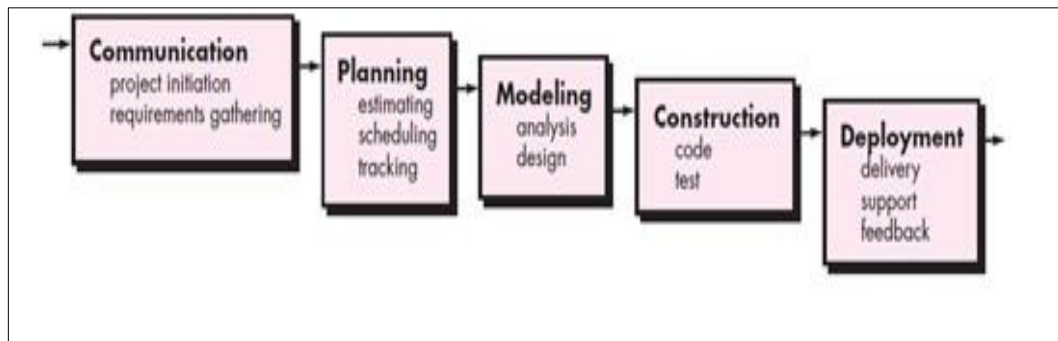
dari bahasa-bahasa pemrograman lain seperti C, C++, Java, PHP, dll. Untuk kepentingan yang spesifik, kode python juga dapat di integrasikan dengan pustaka lain yang ditulis di dalam bahasa C, C++, Java (melalui Jython), dan bahasa-bahasa .NET seperti Visual Basic dan C# (melalui IronPython), (Budi Rahardjo, 2019).

2.9 PostgreSQL

PostgreSQL adalah sebuah sistem basis data yang disebarluaskan secara bebas menurut Perjanjian lisensi BSD. Perangkat lunak ini merupakan salah satu basis data yang paling banyak digunakan saat ini, selain *MySQL* dan *Oracle*. *PostgreSQL* adalah sistem database yang kuat untuk urusan relasi, *open source*. Memiliki lebih dari 15 tahun pengembangan aktif dan sudah terbukti segala rancangan arsitekturnya telah mendapat reputasi tentang “kuat”, “handal”, “integritas data”, dan “akurasi data”, (Regina Obe, Leo Hsu, 2016).

2.10 Waterfall

Model *Waterfall* adalah Model Air Terjun kadang dinamakan siklus hidup klasik (*classic life cycle*), dimana hal ini menyiratkan pendekatan yang sistematis dan berurutan (*sekuensial*) pada pengembangan perangkat lunak. Pengembangan perangkat lunak dimulai dari spesifikasi kebutuhan pengguna dan berlanjut melalui tahapantahapan perencanaan (*planning*), pemodelan (*modeling*), konstruksi (*construction*), serta penyerahan sistem perangkat lunak ke para pelanggan/pengguna (*deployment*), yang diakhiri dengan dukungan berkelanjutan pada perangkat lunak yang dihasilkan (Rohayati, 2014).



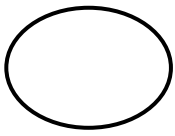
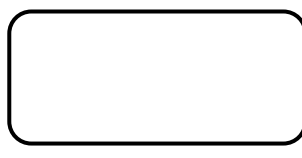
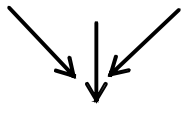
Gambar 2.5 Tahapan *Waterfall*

2.11 Diagram Konteks

Context Diagram adalah gambaran umum tentang suatu sistem yang terdapat didalam suatu organisasi yang memperlihatkan batasan (boundary) sistem, adanya interaksi antara eksternal entity dengan suatu sistem dan informasi secara umum mengalir diantara entity dan sistem. Context Diagram merupakan alat bantu yang digunakan dalam menganalisa sistem yang akan dikembangkan. Simbol-simbol yang digunakan di dalam Context Diagram hampir sama dengan simbol-simbol yang ada pada DFD, hanya saja pada Context Diagram tidak terdapat simbol file (Sukrianto and Oktarina, 2019).

Tabel 2.1 berikut ini adalah Simbol-simbol yang digunakan dalam Diagram Konteks :

Tabel 2.1 Simbol Diagram Konteks

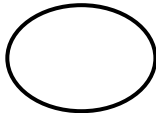
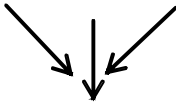
No	Simbol	Keterangan
1	Proses 	Menggambarkan suatu proses atau sistem yang akan dibangun
2	Proses 	Proses dapat digambarkan dengan symbol lingkaran atau persegi panjang dengan sisi-sisi tumpul
3	Entitas/entity 	Menggambarkan entitas atau pengguna dari sistem aplikasi
4	Aliran data (data flow) 	Aliran data yang masuk dan keluar dari sistem.

2.12 Data flow Diagram (DFD)

DFD merupakan gambaran sistem secara logika yang tidak tergantung pada perangkat keras, lunak, struktur data dan organisasi file. Keuntungan dari DFD adalah untuk memudahkan pemakai yang kurang menguasai bidang komputer untuk mengerti sistem yang akan dikerjakan atau dikembangkan (Sukrianto and Oktarina, 2019).

Tabel 2.2 berikut ini adalah Simbol-simbol yang digunakan dalam Data flow Diagram :

Tabel 2.2 Simbol Data Flow Diagram

No	Simbol	Keterangan
1		Proses atau fungsi atau prosedur, pada pemodelan perangkat lunak yang akan diimplementasikan dengan pemrograman terstruktur, maka pemodelan notasi inilah yang seharusnya menjadi fungsi atau prosedur didalam kode program.
2		File atau basis data atau penyimpanan (storage) pada pemodelan perangkat lunak yang akan diimplementasikan dengan pemrograman terstruktur
3		Entitas Luar (external entity) atau masukan (input) atau keluaran (output) atau orang yang memakai/berinteraksi dengan perangkat lunak yang dimodelkan atau sistem lain yang terkait dengan aliran data dari sistem yang dimodelkan.
4		Aliran data merupakan data yang dikirim antar proses, dari penyimpanan keproses, atau dari proses ke masukan (input) atau keluaran (output).

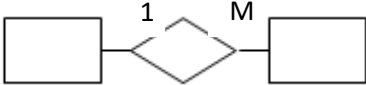
2.13 Relasi Tabel

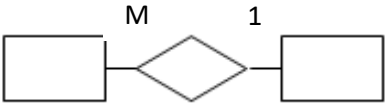
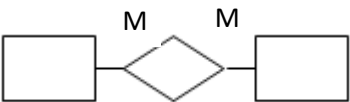
Relational database adalah sebuah kumpulan dari relasi yang telah dinormalisasi dengan nama relasi yang jelas. Relational database merupakan suatu tipe database yang berdasarkan model relational, dimana semua data terlihat oleh user, disusun dalam bentuk tabel-tabel dan semua operasi pada database bekerja pada tabel-tabel tersebut (Cannolly & Begg, 2010). Ada tiga jenis relasi dalam tabel yaitu :

- a) Relasi one-to-one adalah relasi antara satu record dengan satu record dalam tabel lain yang saling berhubungan.
- b) Relasi one-to-many adalah relasi antara satu record dengan lebih dari satu record dalam tabel lain sehingga saling berhubungan.
- c) Relasi many-to-many adalah relasi antara banyak record dengan lebih dari satu record dalam tabel lain yang saling berhubungan.

Tabel 2.3 berikut ini merupakan simbol-simbol yang digunakan dalam relasi tabel :

Tabel 2.3 Simbol Relasi Tabel

Simbol	Keterangan
1. One To One 	Hubungan satu dengan satu
2. One To Many 	Hubungan satu dengan banyak

3. Many To One 	Hubungan banyak dengan satu
4. Many To Many 	Hubungan banyak dengan banyak

2.14 Penelitian Terkait Terdahulu

Tabel 2.4 berikut ini merupakan penelitian terkait dengan penelitian yang sedang dilakukan saat ini :

Tabel 2.4 Penelitian Terkait

No	Nama Penulis	Judul/Tahun Terbit	Uraian
1	Mustakim, Giantika Oktaviani F	Algoritma K-Nearest Neighbor Clasification Sebagai sistem prediksi predikat prestasi mahasiswa Jurnal Sains, Teknologi dan Industri, Vol. 13, No.2, Juni 2016, pp.195 - 202 ISSN 1693-2390 print/ISSN 2407-0939 online	Prestasi Mahasiswa merupakan suatu bentuk dari pencapaian hasil selama mengikuti kegiatan Akademik pada sebuah Perguruan Tinggi. Predikat prestasi mahasiswa diperoleh dari hasil sebuah prediksi. Proses prediksi dilakukan dengan menggunakan metode K-Nearest Neighbor (KNN). Atribut yang digunakan dalam proses prediksi adalah Jenis Kelamin, Jenis Tinggal, Umur, Jumlah Satuan Kredit Semester (SKS), dan Jumlah Nilai Mutu (NM), sehingga dengan menerapkan algortima KNN dapat dilakukan sebuah prediksi berdasarkan kedekatan dari histori data lama

			(training) dengan data baru (testing).
	Pengembangan yang dilakukan		Penelitian ini dilakukan menggunakan metode yang sama yaitu metode <i>K-Nearest Neighbor</i> . Pada penelitian ini terdapat 8 variabel yang digunakan yaitu batuk > 3 minggu, batuk berdarah, batuk berdarah, sesak nafas, nyeri dada, mengi, dahak kental berwarna hijau, batuk menetap dan timbul berulang.
2	Ketut Artaye	Penerapan Metode Klasifikasi Nave Bayes Pada Prediksi Waktu Kelulusan Beasiswa Ibi Darmajaya International Conferences on Information Technology and Business (ICITB), 20th -21st August 201	Penerapan Metode Klasifikasi Nave Bayes Pada Prediksi Waktu Kelulusan Beasiswa Ibi Darmajaya, Penelitian ini menggunakan nave bayes, metode Classifier untuk memprediksi waktu kelulusan yang dapat menentukan ketepatan waktu belajar dalam jurusan teknik informatika. Penelitian ini menggunakan lima atribut yang digunakan.
	Pengembangan yang dilakukan		Penelitian ini menggunakan K-nn sebagai metode yang digunakan untuk mencari nilai kedekatan. Variabel yang digunakan pada penelitian ini sebanyak 8 variabel yaitu batuk > 3 minggu, batuk berdarah, batuk berdarah, sesak nafas, nyeri dada, mengi, dahak kental berwarna hijau, batuk menetap dan timbul berulang.
3	Andi Maulida Argina	Penerapan Metode Klasifikasi K-Nearest	Diabetes adalah suatu penyakit metabolik yang diakibatkan oleh

		Neighbor pada Dataset Penderita Penyakit Diabetes Indonesian Journal of Data and Science Vol 1, No 2, Juli 2020, pp. 29-33	meningkatnya kadar glukosa atau gula darah. Gula darah sangat vital bagi kesehatan karena merupakan sumber energi yang penting bagi sel-sel dan jaringan. Penelitian dilakukan dengan melakukan pembagian data training dan data testing, data yang digunakan sebanyak 77 data, dengan pembagian sebesar 90% sebagai data training dan 10% sebagai data tesing. Tahapan selanjutnya adalah menerapkan metode KNN
	Pengembangan yang dilakukan		Penelitian yang dilakukan menggunakan metode K-nn dimana Variabel yang digunakan pada penelitian ini sebanyak 8 variabel yaitu batuk > 3 minggu, batuk berdahak, batuk berdarah, sesak nafas, nyeri dada, mengi, dahak kental berwarna hijau, batuk menetap dan timbul berulang. Data yang digunakan sebanyak 132 data
4	Mei Lestari	Penerapan Algoritma Klasifikasi Nearest Neighbor (K-Nn) Untuk Mendeteksi Penyakit Jantung Faktor Exacta 7(4): 366-371, 2014 ISSN: 1979-276X	Jantung merupakan organ manusia yang berperan dalam Sistem peredaran darah. Pada penelitian ini digunakan 110 records pasien. 100 records digunakan sebagai data latih (training data) dan 10 records digunakan sebagai data uji (testing data).
	Pengembangan yang dilakukan		Penelitian ini dilakukan menggunakan metode yang sama yaitu metode

			<i>K-Nearst Neighbor</i> .Pada penelitian ini menggunakan 132 data sebagai data training dan 10 untuk data testing.
5	Septilia Arfida, Maruli Tua Siahaan	Rancang Bangun Sistem Pendiagnosa Penyakit Osteoporosis Pada Wanita Menggunakan Metode Fuzzy Inference Sistem Tsukamoto Jurnal Informatika, Vol. 16, No. 1, Juni 2016	Osteoporosis adalah tulang yang keropos, yaitu penyakit yang ditandai dengan berkurangnya massa tulang rendah, disertai gangguan mikro-arsitektur tulang dan penurunan kualitas jaringan tulang. Osteoporosis dapat terdeteksi dengan alat khusus melalui pemeriksaan radiologi. penyakit Osteoporosis pada wanita menggunakan metode Fuzzy Inference System Tsukamoto. Metode ini dipilih untuk melakukan diagnosa penyakit osteoporosis yang memiliki 5 variabel yaitu Usia, Aktifitas, Kalsium, Tinggi Badan dan Berat Badan.
	Pengembangan yang dilakukan		Penelitian yang dilakukan menggunakan metode K-nn dimana Variabel yang digunakan pada penelitian ini sebanyak 8 variabel yaitu batuk > 3 minggu, batuk berdahak, batuk berdarah, sesak nafas, nyeri dada, mengi, dahak kental berwarna hijau, batuk menetap dan timbul berulang. Data yang digunakan sebanyak 132 data.