

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/321804055>

DATA MINING K-Nearest Neighbor

Book · December 2017

CITATIONS

0

READS

1,803

3 authors, including:



Risky Aswi Ramadhani

Universitas Nusantara PGRI Kediri

7 PUBLICATIONS 3 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



PENENTUAN PENERIMA ZAKAT DENGAN METODE FUZZY [View project](#)

DATA MINING



K-Nearest Neighbor (KNN)

Risa Helilintar, Risky Aswi Ramadhani, Siti Rochana

DATA MINING

K-Nearest Neighbor

**Kediri; Fakultas Teknik Universitas Nusantara PGRI
Kediri,**

2017 ;52 hlm.;15,5x23 cm

ISBN:

DATA MINING

K-Nearest Neighbor

Penulis

Risa Helilintar, M.Kom

Risky Aswi Ramadhani, M.Kom

Siti Rochana, M.Pd

Penyunting

Danar Putra Pamungkas, M.Kom

Cover Ilustrai

Risky Aswi Ramadhani, M.Kom

@2017Cetakan Pertama

Diperbolehkan mengutip sebagian atau seluruh isi buku ini dengan cara apapun termasuk dengan menggunakan mesin fotocopy, dengan atau tanpa izin tertulis dari penulis.

PENERBIT

Fakultas Teknik Universitas Nusantara PGRI Kediri

Kampus II, Mojoroto Gang I No.6 Kediri

Email : ft@unpkediri.ac.id

**DATA MINING
K-NEAREST NEIGHBOR**



**Risa Helilintar, M.Kom
Risky Aswi Ramadhani, M.Kom
Siti Rochana, M.Pd**

**FAKULTAS TEKNIK
UNIVERSITAS NUSANTARA PGRI KEDIRI**

KATA PENGANTAR

Buku Data Mining Membahas Konsep dan Aplikasinya 1 ini membantu Mahasiswa Untuk Mempelajari data mining dan aplikasinya. Buku ini disusun dengan menggunakan bahasa yang mudah dipahami.

Buku ini terdiri dari beberapa 9 bab yaitu

BAB I Pengertian Data Mining

BAB II Data

BAB III Proses *Knowledge Data Discovery*

BAB IV Manfaat Data Mining

BAB V Teknik – teknik data mining

BAB VI Hepatitis

BAB VII KNN

BAB VIII *LEARNING*

BAB IX Pemanfaatan Di Penelitian

Dengan terbitnya buku ini diharapkan proses belajar mengajar di Universitas Nusanatara PGRI Kediri, Fakultas Teknik, Prodi Teknik Informatika dapat meningkat. dan membantu mahasiswa mempelajari data Mining

Buku ini merupakan karya pertama, penulis masih perlu banyak belajar tentang kedalaman materi, cara penulisan, jika ada kesalahan dalam penulisan ini penulis mohon maaf sebesar-besarnya.

DAFTAR ISI

Sampul Dalam	iii
KATA PENGANTAR.....	iv
DAFTAR ISI.....	v
BAB I PENGERTIAN DATA MINING	1
1.1 Data Where House.....	1
1.2 Permasalahan dalam data Mining.....	6
1.3 Teknik Data Mining	8
BAB II DATA.....	11
2.1 Operasi Dasar Basisi Data	13
2.2 Penerapan Basis data	15
BAB III PROSES KNOWLEDGE DATA	
DISCOVERY.....	17
3.1 Data Selection.....	17
3.2 Pre-processing/ Cleaning.....	18
3.3 Transformation	18
BAB IV MANFAAT DATA MINING	25
4.1 Segmentasi Pasar	25
4.2 Analisis keranjang penjualan.....	27
4.3 Analisis kecenderungan.....	27
4.4 Intelligence Marketing.....	27
BAB V TEKNIK – TEKNIK DATA MINING	29
5.1 Association Rules	29
5.2 Clustering	29
5.3 Deskripsi	30
5.4 Estimasi	30
5.5 Prediksi	31

BAB VI HEPATITIS	32
6.1 Hepatitis A.....	32
6.2 Hepatitis B	33
BAB VII KNN	36
BAB VIII LEARNING	39
8.1 Belajar Machine Learning	39
8.2 Aplikasi Machine Learning	40
8.3 Dampak Machine Learning di Masyarakat	42
BAB IX PEMANFAATAN DI PENELITIAN	44
9.1 K-Nearest Neighbour (K-NN).....	44
9.2 Variabel	44
9.3 Data Training	45
9.4 Data Testing.....	47
9.5 DFD Level Kontek	48
9.6 Prototipe	49
9.6.1 Tampilan Data Training	49
9.6.2 Tampilan Data Latih (mendiagnosa).....	49
9.6.3 Hasil Diagnosa	50
DAFTAR PUSTAKA.....	51

BAB I

PENGERTIAN DATA MINING

Data Mining adalah Serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basisdata dengan melakukan penggalian pola-pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basisdata

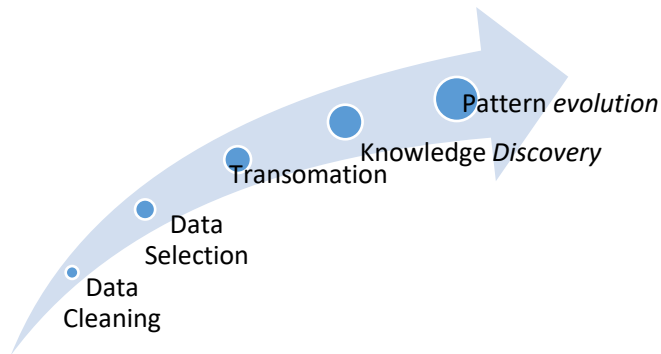
Data Mining merupakan analisis dari peminjaman Kumplan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data(Larose, 2005)

Data Minining biasa juga dikenal nama lain seperti : *Knowledge discovery (mining) in databases (KDD)*, ekstraksi pengetahuan (*knowledge extraction*) Analisa data/pola dan kecerdasan bisnis (*business intelligence*) dan merupakan alat yang penting untuk memanipulasi data untuk penyajian informasi sesuai kebutuhan user dengan tujuan untuk membantu dalam analisis koleksi pengamatan

perilaku, secara umum definisi data-mining dapat diartikan sebagai berikut

- a. Proses penemuan pola yang menarik dari data yang tersimpan dalam jumlah besar.
- b. *Ekstraksi* dari suatu informasi yang berguna atau menarik (non-trivial, implisit, sebelumnya belum diketahui potensial kegunaannya) pola atau pengetahuan dari data yang disimpan dalam jumlah besar.
- c. *Ekplorasi* dari analisa secara otomatis atau semi otomatis terhadap data-data dalam jumlah besar untuk mencari pola dan aturan yang berarti.

Pada saat ini data mining menarik perhatian para pelaku usaha, karena pelaku usaha ingin memanfaatkan data mining untuk meningkatkan produktivitas usaha-nya, selain itu data mining juga dapat digunakan dibidang kesehatan anatara lain mendiagnosa penyakit, mengelompokan kesehatan masyarakat, dll. Untuk memanfaatkan data mining agar optimal perlu adanya beberapa proses, berikut ini adalah proses-proses pengolahan data mining



1. *Data cleaning* (untuk menghilangkan noise data yang tidak konsisten) *Data integration* (di mana sumber data yang terpecah dapat disatukan). Biasanya terdapat data yang kurang bagus untuk dimasukkan dalam kelengkapan data perusahaan karena hanya akan dianggap tidak valid bahkan untuk data yang hilang. Sehingga data yang seperti itu lebih baik dibuang. Cleaning data merupakan hal yang sangat penting karena cleaning dapat mempercepat proses query. Pembersihan data juga digunakan untuk membersihkan redundansi data, yang di maksud redundansi data adalah pengulangan data lebih dari satu
2. *Data selection* (di mana data yang relevan dengan tugas analisis dikembalikan ke dalam database)
3. *Data transformation* (di mana data berubah atau bersatu menjadi bentuk yang tepat untuk menambang dengan ringkasan performa atau operasi agresif) misalkan perubahan file bulan disesuaikan dengan kebutuhan.

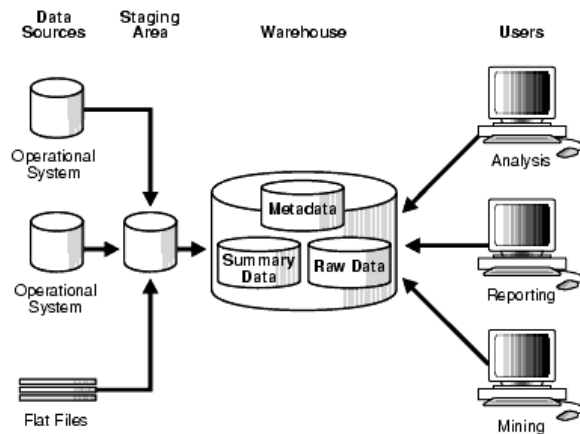
4. *Knowledge Discovery* (proses esensial di mana metode yang intelegen digunakan untuk mengekstrak pola data)
5. *Pattern evolution* (untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan atas beberapa tindakan yang menarik)

1.1 Data Where House

Data Warehouse adalah sumber data yang terdiri dari data masa lalu dan data saat ini sumber data data berasal dari transaksi, baik itu transaksi internal maupun external., data ini akan digunakan sebagai bahan untuk menganalisa atau meperdiksi.pada saat ini pemanfaatan datawarehouse sudah dimanfaatkan secara optimal (Hermawati,2013).

Pada saat ini datawarehouse menjadi perhatian penting bagi manajemen suatu instansi atau organisasi, karena dengan menganalisa data where house suatu intasi atau organisasi bisa melihat langkah yang akan dilakukan beberapa tahun kedepan. Data Warehouse juga digunakan untuk pengambilan keputusan suatu instansi atau organisasi, dengan catatan untuk mengambil keputusan tersebut diambil dari data internal dan external.

Data Where house merupakan sebuah data yang digunakan untuk mendukung pengambilan keputusan, bukan untuk melaksanakan proses transaksi. Pengembalian data where house akan disimpan dan dianalisa sehingga membentuk sebuah pola.berikut ini adalah skema datawarehouse



Gambar 1.1. Skema data Where House
(<https://docs.oracle.com>)

Data source merupakan data yang belum diolah, data ini belum terpola, jadi data ini masih bersifat abstrak. Data source diambil dari data transaksi. Setelah data source terkumpul dalam beberapa waktu data tersebut diolah pada proses metadata, Summary data, dan raw data. Setelah itu data analisis, data rporting , dan data mining bertugas mengambil keputusan dari data meta.

1.2 Permasalahan dalam data Mining

Sistem data basis data adalah sebuah sistem(tatanan) yang digunakan untuk mendukung sebuah proses tertentu(Fahtasyah, 2012). Sistem Basis data merupakan serangkaian tabel yang befungsi pada satu tujuan, misalkan untuk membuat sebuah sistem yang berkaitan dengan rekam medis pasien di butuhkan 3 tabel dasar yaitu tabel biodata pasien, tabel penyakit, dan obat.

Sistem Basis data merupakan sebuah objek yang pasif pada sistem ini user tidak dapat mempergunakan sistem secara jauh karena sistem ini hanya bertugas sbagai penyimpanan. Operasi yang biasanya terjadi pada sistem basis data adalah create, update, delete. Sistem basis data biasanya disimpan pada sebuah server. Lebih lengkapnya sistem basis data terdiri dari beberapa komponen yaitu

- *Hardware(Sever)*
- Sistem Operasi(Windos, Linux, Mac)
- Basis data(My Sql,Acses, Oracle)
- Sistem(Aplikasi Pengelola Perangkat Lunak
- Pemakai(Admin, Owner)
- Aplikasi(Sebuah aplikasi perangkat lunak yang berfungsi untuk mengoptimalkan basis data)

Yang menjadi permasalahan sistem basis data adalah saat pertama kali di develop sistem hanya berfungsi sebagai alat untuk mendukung pekerjaan, mengurangi tenaga kerja. Belum ada tujuan yang jelas bahwa sistem basis data tersebut akan diolah dan menghasilkan suatu pengetahuan.

Untuk mengolah sistem basis data agar dapat diambil pengetahuannya perlu diadakan beberapa tahapan proses seperti cleaning, normalisasi summarizing. Selain itu pada data base sering terjadi kasus hilangnya data karena terjadi crash atau pada saat input ada beberapa data yang tidak terproses. Untuk mengatasi permasalahan tersebut ada beberapa cara yang bisa dilakukan antara lain.

- Mengabaikan nilai-nilai yang hilang
- Menghilangkan record yang berhubungan
- Menembak nilai yang hilang dari nilai yang diketahui
- Memperlakukan data yang hilang sebagai sebuah nilai khusus yang dimasukkan sebagai tambahan dalam domain atribut.
- Atau menghitung nilai rata-rata yang hilang menggunakan nilai Bayesian.

Data yang mengandung noise cenderung sulit diolah dengan metode-metode statistik seperti Gaussian. Agar data tersebut bisa diolah dengan Gaussian maka perlu dilakukan proses cleaning.

1.3 Teknik Data Mining

Ada banyak teknik data mining setiap teknik memiliki fungsi dan karakter tersendiri. Tergantung permasalahan dan tujuan yang ingin dicapai dengan data mining tersebut. Berikut ini adalah macam-macam teknik data mining Klasifikasi, Clustering, Association Rule Discovery, dan Regression, Deviation Detection.

Buku ini akan membahas Klasifikasi, klasifikasi digunakan untuk menentukan sebuah record baru ke salah satu data yang sudah didefinisikan sebelumnya, kelas ini disebut dengan data training. Metode-metode yang menggunakan klasifikasi data adalah *decision tree*, *K- nearest neighbor*, dan *naïve bayes*

Klasifikasi data adalah sebuah pencocokan data baru (data testing) dengan data training, dari hasil pencocokan tersebut akan memunculkan sebuah keputusan dari hasil yang ditargetkan dari data training. Berikut ini adalah contoh tabel yang

digunakan untuk mencocokkan data training dengan data testing.



Gambar 1.2. Data traing dan data testing(Bertalya,2009)

Data training adalah data yang diambil dari masa lampau, tujuan diambilnya data ini adalah sebagai sumber pengetahuan yang akan muncul lagi pada priode mendatang. Jadi jika pada data traing sudah ada beberapa karakter yang tersimpan, saat data testing diipukan dan data tersebut sesuai daengan data training kemungkinan besar data testing tersebut memiliki kesaaman sifat denngan data training.

Semakin banyak jumlah data training maka sebuah keputusan yang diambil oleh data mining semakin akurat. Klasifikasi juga dapat meningkatkan kemampuan tetang suatu hal dengan bertambahnya

data di data training, kemampuan belajar dari metode ini disebut dengan learning.

BAB II

DATA

Data merupakan sesuatu angka atau huruf yang belum memiliki arti dan manfaat, perlu beberapa tahapan untuk mengolah data menjadi sesuatu yang berguna bagi khalayak umum, data yang berguna pada khalayak umum disebut dengan informasi. Data disebut juga sebagai representasi dunia nyata misalnya pada manusia ada data alamat, ukuran sepatu. Sebagai kesatuan istilah data dapat diartikan dari berbagai sudut pandang antara lain:

- Himpunan kelompok data dikelompokkan sedemikian rupa dengan folder agar dapat dimanfaatkan dengan mudah, pengelompokan data digital maupun data fisik harus sesuai dengan nama atau penanggalan. Tujuannya agar saat membutuhkan dapat dicari dengan mudah.
- Data merupakan suatu relasi yang saling terhubung satu sama lain data tidak diperbolehkan mengalami redudansi(pengulangan), Id data harus bersifat unik tidak boleh sama dengan yang lainnya. Selain itu pembuatan tabel pada database harus sesuai dengan kebutuhan. Tujuannya agar dapat menghemat stroge hardisk.

- Penyimpanan data pada elektronik disebut dengan data base, data base terdiri dari beberapa tabel, sedangkan tabel terdiri dari beberapa record.



Gambar 2.1. Struktur database

Data Eletronik atau data base memiliki struktur yang lengkap mulai dari data base, tabel, field dan record

Basis data adalah sebuah lemari arsip eletronik yang besar, jika dalam penamaan tabel data base tidak sesuai dengan ketentuan, saat mengolah dengan data mining biasanya mengalami I beberapa kendala seperti tidak adanya kesepahaman antar programmer.berikut ini adalah pembuatan tabel database yang terstruktur

wp_comments	Browse	Structure	Search	Insert	Empty	Drop	1	InnoDB	utf8mb4_unicode_520_ci	96	KiB	-
wp_links	Browse	Structure	Search	Insert	Empty	Drop	0	InnoDB	utf8mb4_unicode_520_ci	32	KiB	-
wp_options	Browse	Structure	Search	Insert	Empty	Drop	129	InnoDB	utf8mb4_unicode_520_ci	528	KiB	-
wp_postmeta	Browse	Structure	Search	Insert	Empty	Drop	1	InnoDB	utf8mb4_unicode_520_ci	48	KiB	-
wp_posts	Browse	Structure	Search	Insert	Empty	Drop	3	InnoDB	utf8mb4_unicode_520_ci	80	KiB	-
wp_termmeta	Browse	Structure	Search	Insert	Empty	Drop	0	InnoDB	utf8mb4_unicode_520_ci	48	KiB	-
wp_terms	Browse	Structure	Search	Insert	Empty	Drop	1	InnoDB	utf8mb4_unicode_520_ci	48	KiB	-
wp_term_relationships	Browse	Structure	Search	Insert	Empty	Drop	1	InnoDB	utf8mb4_unicode_520_ci	32	KiB	-

Gambar 2.2. Struktur tabel CMS Wordpress

Nama tabel pada database cms wordpress tersebut memiliki awalan WP_ tujuannya adalah saat dilakukan pengerjaan oleh beberapa progroamer, progemer yang tidak membuat databse tidak mengalami kesulitan.

2.1 Operasi Dasar Basisi Data

Basis data adalah sebuah data yang berkaitan dengan tabel field record, database yang dibuat harus memiliki kecepatan, akurasi, dan efektif. Agar ketiga unsur tersebut bisa tecapai maka dalam pembuatan database perlu diperhatikan strutur tabel dan relasi yang ada didalamnya. Untuk membuuat sebuah database ada beberapa operasi yang berkenaan dengan database diantaranya

- Pembuatan basis data baru (*create database*), yang identik dengan pembuatan lemari arsip yang baru.
- Penghapusan basis data (*drop database*), yang identik dengan perusakan lemari arsip (sekaligus beserta isinya jika ada).
- Pembuatan file/tabel baru ke suatu basis data (*create table*), yang identik dengan penambahan map arsip baru ke sebuah lemari arsip yang telah ada.

- Penghapusan file/tabel dari suatu basis data (*drop table*), yang identik dengan perusakan map arsip lama yang ada di sebuah lemari arsip.
- Penambahan/pengisian data baru ke sebuah file/tabel di sebuah basis data (*insert*), yang identik dengan penambahan lembaran arsip ke sebuah map arsip.
- Pengambilan data dari sebuah file/tabel (*retrieve/search*), yang identik dengan pencarian lembaran arsip dari sebuah map arsip.
- Pengubahan data dari sebuah file/tabel (*update*), yang identik dengan perbaikan isi lembaran arsip yang ada di sebuah map arsip.
- Penghapusan data dari sebuah file/tabel (*delete*), yang identik dengan penghapusan sebuah lembaran arsip yang ada di sebuah map arsip.

Operasi dalam pembuatan tabel biasanya dilakukan sekali saja sedangkan operasi yang berhubungan dengan insert data biasanya dilakukan rutin.

Pada pembuatan laporan biasanya terdapat beberapa penggunaan fungsi logika seperti if, else, tujuannya agar saat menampilkan data, data yang tampil sesuai dengan kebutuhan. Misalkan

penggunaan query yang memiliki kondisi khusus untuk menampilkan data

SELECT * FROM `data` where status='no'

id	muntah	demam	mata	bak	badan	nafsu	perut	status
1	1	2	1	2	2	2	1	No
3	2	2	2	2	1	1	1	No

Gambar 2.3. Contoh menampilkan data pasien yang berstatus tidak terjangkit

Pada Gambar diatas ditampilkan where ='no' permintaan query diatas adalah menampilkan data pasien yang tidak terjangkit hepatitis.

2.2 Penerapan Basis data

Hampir semua instansi menggunakan basis data dalam kegiatannya sehari-hari, tujuan penggunaan basis data adalah untuk mengoptimalkan kinerja, menghemat waktu, dan mengurangi SDM. Berikut ini adalah contoh penggunaan Basis data.

- Kepagawaian
- Pergudangan
- Akuntansi
- Reservasi
- Layanan Pelanggan
- DLL

Sedangkan bentuk organisasi yang menggunakan basisdata adalah

- Rumah sakit

Untuk mengolah data pasien, rekam medis pasien, data kepegawaian, laporan-laporan, pengananaan pembayaran pasien.

- Perbankan

Diperbankan basisdata digunakan untuk pengelolaan pinjaman, pengelolaan tabungan, data nasabah pelayanan Informasi

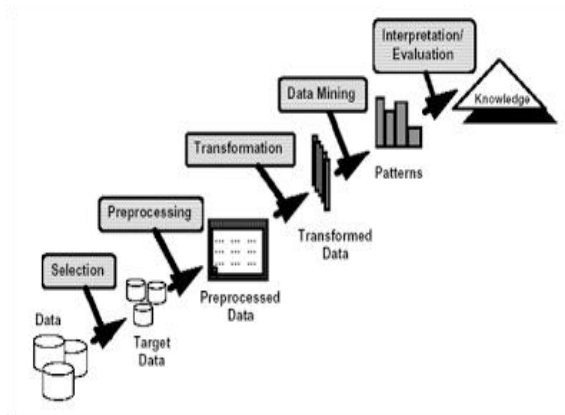
- Pendidikan Sekolah

Pada dunia pendidikan basisdata digunakan untuk penyimpanan nilai, dana bos siswa, pendaftaran siswa baru, gaji guru

BAB III

PROSES *KNOWLEDGE DATA DISCOVERY*

Knowledge Data Discovery Berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah kumpulan data. *Knowledge discovery in databases* (KDD) adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti oleh user, yang di sajikan dalam bentuk grafik maupun tabel



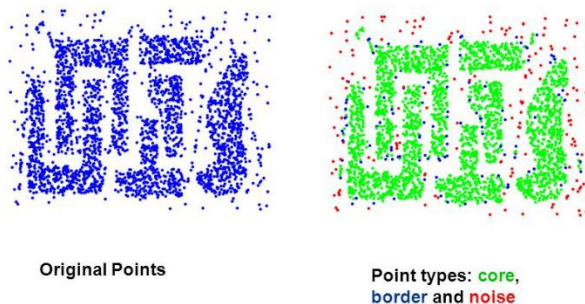
Gambar 3.1 KDD

3.1 Data Selection

Data Selection adalah proses memnentukan data target , pemilihan data target ini memfokuskan pada variabel atau data yang akan digunkanan. Pada tahapan ini akan dilakukan pemilihan variabel yang tepat sehingga data yang dipilih dapat beroperasi sesuai dengan fungsinya.

3.2 Pre-processing/ Cleaning

Cleaning data harus memenuhi beberapa persyaratan , yaitu data harus terbebas dari redudesi, data harus konsisten, dan tidak adanya noise saat K-DD(Rahm).



Gambar. 3.2 Noise data

Pada Gmbar diatas dijelaskan bahawa data yang bersifat tidak teratur bisa dibersihkan atau diapus, selain itu pada beberapa kasus dta tersebut diubah pada proses tranformasi.

3.3 *Transformation*

Proses ini adalah proses indentifikasi fitur-fitur yang akan digunakan misalkan toolsnya, atau metode pengambilan datanya, pada tahapan ini harus melihat tujuan yang akan dicapai terlebih dahulu

Dalam data *transformation*, terdapat beberapa pendekatan/teknik untuk melakukan transformasi data, yaitu *smoothing*, *generalization*, *normalization*, *aggregation*, dan *attribute construction*(Hartanto, 2011)

a. *Smoothing*

Smoothing dilakukan jika data mengandung noise/nilai yang tidak valid terhadap data yang di-mining. Untuk mengatasinya harus dilakukan *smoothing* (dengan memperhatikan nilai-nilai tetangga). Berikut teknik atau metode untuk *smoothing*:

- *Binning* Metode binning dilakukan dengan memeriksa “nilai tetangga”, yaitu nilai-nilai yang ada disekelilingnya. Berikut adalah langkah-langkah metode binning: 1. Data diurutkan dari yang terkecil sampai dengan yang terbesar. 2. Data yang sudah urut

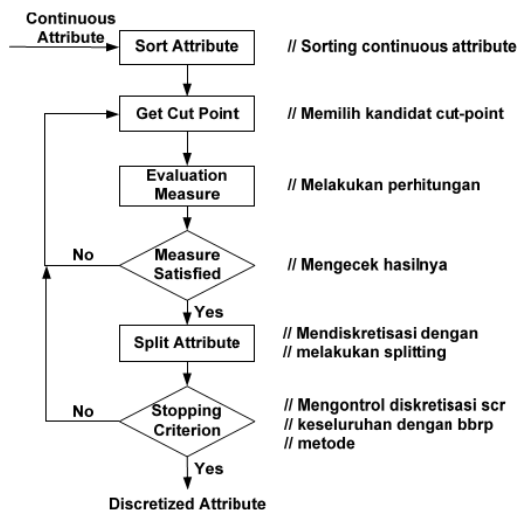
kemudian dipartisi ke dalam beberapa bin. Teknik partisi ke dalam bin ada 2 (dua) cara: *equal-width (distance) partitioning* dan *equaldepth (frequency) partitioning*. 3. Dilakukan *smoothing* dengan tiga macam teknik, yaitu: *smoothing by binmeans*, *smoothing by bin-medians*, dan *smoothing by bin-boundaries*.

- *Clustering* Digunakan untuk menyingkirkan *outliers* (keluar jauh-jauh dari cluster/*centroid*), data yang memiliki noise. Algoritma *k*-Means yang merupakan kategori metode partitioning dapat digunakan jika ukuran *database* tidak terlalu besar. Algoritma ini didasarkan pada nilai tengah dari objek yang ada dalam cluster. Algoritma *k*-Means meminta inputan parameter *k*, dan mempartisi satu set *n* objek ke dalam *k* cluster sehingga menghasilkan tingkat kemiripan yang tinggi antar objek dalam kelas yang sama (*intra-class similarity*) dan tingkat kemiripan yang paling rendah antar objek dalam kelas yang berbeda (*inter-class similarity*). Kemiripan

cluster diukur dengan menghitung nilai tengah dari objek yang ada di dalam cluster.

b. Generalization

Generalization atau generalisasi adalah ketika data level rendah (*low-level data*) diganti dengan konsep yang lebih tinggi, yaitu dengan melakukan *diskretisasi*. Teknik diskretisasi dapat digunakan untuk mereduksi sekumpulan nilai yang terdapat pada atribut continuous, dengan membagi *range* dari atribut ke dalam *interval*.



Gambar 3.3. Proses Diskretisasi

Gambar 3.3. Proses Diskretisasi

Proses diskretisasi secara umum terdiri dari 4 tahapan (gambar 2), yaitu: 1. *Sorting*, melakukan sorting nilai atribut continuous yang mau didiskretisasi. 2. *Memilih “cut-point”*, banyak fungsi evaluasi yang dapat digunakan seperti binning dan pengukuran entropy. 3. *Splitting*, dilakukan evaluasi cut-point yang ada dan pilih satu yang terbaik dan lakukan split range nilai atribut continuous ke dalam dua partisi. Diskretisasi berlanjut untuk tiap partisi sampai kondisi berhenti tercapai. 4. *Stopping criterion*, diperlukan untuk menghentikan proses diskretisasi. Ada 5 metode untuk melakukan diskretisasi pada atribut continuous, yaitu: binning, cluster analysis, *histogram analysis*, *entropy-based discretization*, dan *segmentation by “natural partitioning”*. Dua metode pertama telah dibahas pada data smoothing, pada subbab ini akan dibahas 3 metode yang lainnya.

1. **Histogram Analysis** Seperti binning sebelumnya, pertama data harus diurutkan dahulu kemudian membagi data ke dalam keranjang dan menyimpan nilai rata-rata

(total) tiap keranjang. Untuk menentukan jumlah keranjang dan nilai atribut yang dipartisi, ada beberapa aturan partisi yaitu: *equal-width*, *equal-depth*, *V-Optimal*, dan *MaxDiff*. V-Optimal dan MaxDiff histogram cenderung lebih akurat dan praktis(Hartanto, 2011).

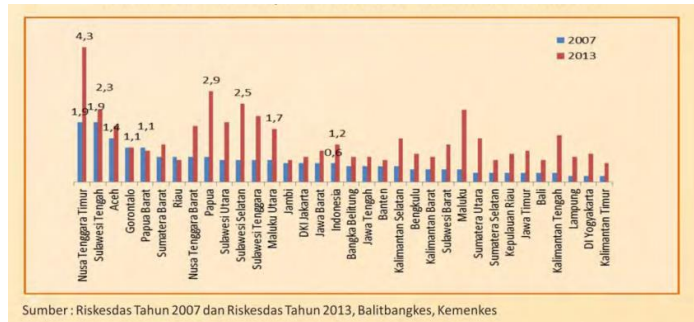
2. Entropy-Based Discretization Diskretisasi berdasarkan nilai entropy merupakan metode diskretisasi secara *supervised*. Seperti metode diskretisasi lainnya, atribut yang mau didiskretisasi diurutkan dahulu. Algoritma *supervised* ini menggunakan *class information entropy* dari partisi untuk memilih batas bin dalam melakukan diskretisasi.

3. Data mining

Pada proses ini adalah proses penentuan pola yang akan diambil, selain itu pada proses ini harus ditentukan metode yang cocok untuk memecahkan permasalahan yang didapat misanya pemilihan metode K-NN untuk mendiagnosa penderita Hepatitis.

4. Interpretation/ Evaluation

Adalah sebuah proses dimana seorang analis bertugas menampilkan hasil dari proses mining, cara menampilkan data, tampilan data ini harus berupa diagram atau tabel yang mudah dipahami oleh kalayak umum, baik itu dari user yang memiliki latar belakang IT maupun disiplin ilmu lainnya, berikut ini adalah cohtoh diagram.



Tabel 1. Penderita Hepatitis

kerja karena pekerjaan telah digantikan oleh alat teknologi *machine learning* adalah suatu permasalahan yang harus dihadapi. Ditambah dengan ketergantungan terhadap teknologi akan semakin terasa. Manusia akan lebih terlena oleh kemampuan gadget-nya sehingga lupa belajar untuk melakukan suatu aktivitas tanpa bantuan teknologi.

BAB IX

PEMANFAATAN DI PENELITIAN

9.1 K-Nearest Neighbour (K-NN)

Prinsip kerja *k-Nearest Neighbor (k-NN)* adalah mencari jarak terdekat antara data yang akan dievaluasi dengan *k* tetangga (*Neighbor*) terdekatnya dalam data pelatihan. Berikut urutan proses kerja *k-NN* (Gorunescu, 2011):

1. Menentukan parameter *k* (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak *euclidean (euclidean distance)* masing-masing obyek terhadap data sampel yang diberikan.

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

Keterangan:

x_1 = Sampel Data

x_2 = Data Uji / Testing

i = Variabel Data

d = Jarak

p = Dimensi Data

9.2 Variabel

Dari hasil observasi maka ditemukan bahwa hepatitis dapat dikenali dengan beberapa hal yaitu muntah minimal 3 kali, demam 3 hari berturut-turut,

warna mata kuning, BAK kuning teh, badan lemas, nafsu makan menurun, nyeri perut atas.

Dari gejala hasil observasi maka akan digunakan metode K-NN untuk mengenali apakah orang tersebut menderita Hepatitis.

9.3 Data Training

Berikut ini adalah data training penderita hepatitis, bisa dilihat pada tabel 1.

G1=Muntah minimal 3 kali

G5=Badan Lemas

G2=Demam 3 hari berturut-turut

G6=Nasus Makan Menurun

G3=Warna Mata Kuning

G7=Nyeri Perut atas

G4=BAK Kuning Teh

Tabel 9.1. Penderita dan gejala

No	G1	G2	G3	G4	G5	G6	G7	Keterangan
231	Ya	Ya	Ya	Ya	Ya	Ya	Ya	Terjangkit
232	Ya	Ya	Tidak	Ya	Ya	Ya	Tidak	Tidak terjangkit
233	Ya	Ya	Ya	Ya	Ya	Ya	Tidak	Terjangkit
234	Ya	Ya	Ya	Ya	Ya	Ya	Ya	Tidak terjangkit
235	ya	ya	Ya	Ya	Ya	Ya	Ya	terjangkit

Nilai Kedekatan

- a. Muntah 3 kali

Tabel 9.2. Gejala Muntah 3 kali

Muntah minimal 3 kali	ya	tidak
ya	1	0
tidak	0	1

- b. Demam 3 Hari

Tabel 9.3. Gejala Demam 3 Hari

Muntah minimal 3 kali	ya	tidak
ya	1	0
tidak	0	1

- c. Warna Mata Kuning

Tabel 9.4. Warna Mata Kuning

Warna Mata Kuning	ya	tidak
ya	1	0
tidak	0	1

- d. BAK Kuning Teh

Tabel 9.5. BAK Kuning Teh

Muntah minimal 3 kali	ya	tidak
ya	1	0
tidak	0	1

e. Badan Lemas

Tabel 9.6. Badan Lemas

Muntah minimal 3 kali	ya	tidak
ya	1	0
tidak	0	1

f. Nafsu Makan Menurun

Tabel 9.7. Nafsu Makan Menurun

Muntah minimal 3 kali	ya	tidak
ya	1	0
tidak	0	1

g. Nyeri Perut atas

Tabel 9.8. Gejala Nyeri Perut atas

Muntah minimal 3 kali	ya	tidak
ya	1	0
tidak	0	1

9.4 Data Testing

Data Testing adalah data dari pasien yang sudah diambil gejala-gejalanya, data testing akan dicocokkan dengan data training untuk mencari data terdekatnya, berikut ini adalah contoh dari data testing

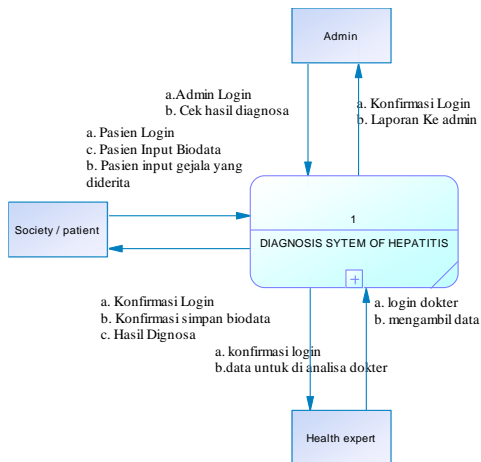
Tabel 9.9. Penderita dan gejala

No	G1	G2	G3	G4	G5	G6	G7	Keterangan
236	Ya	Ya	Ya	Tidak	Ya	Tidak	Tidak	?

Dari hasil tes data testing maka, ditemukan bahwa pasien dengan ID 236 dinyatakan terjangkit, karena memiliki kedekatan dengan pasien no 241 yang berstatus terjangkit, kedekatan pasien tersebut adalah 0,83.

9.5 DFD Level Kontek

Berikut ini adalah diagram level Kontek untuk Perancangan Sistem Diagnosa Penyakit Hepatitis Menggunakan Metode KNN. DFD ini terdiri dari 3 entitas yaitu admin, pasien dan Dokter

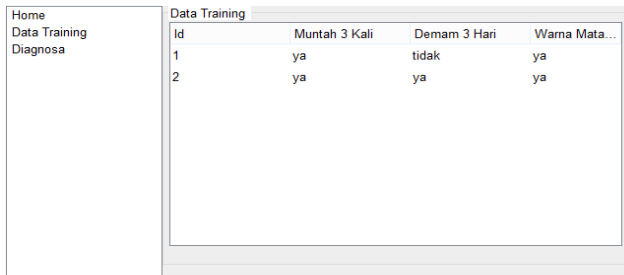


Gambar 9.1 DFD Level Kontex

9.6 *Prototipe*

Berikut ini adalah Prototipe Perancangan Sistem Diagonosa Penyakit Hepatitis Menggunakan Metode KNN.

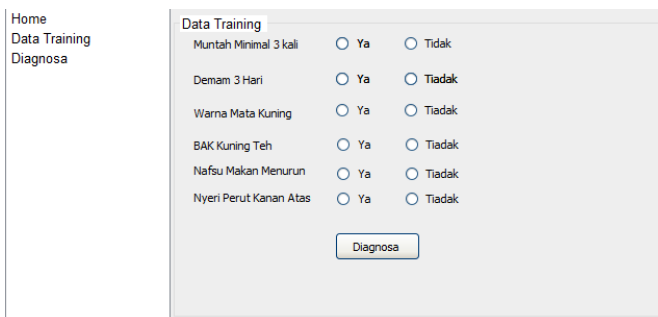
9.6.1 Tampilan Data Training



Id	Muntah 3 Kali	Demam 3 Hari	Warna Mata...
1	ya	tidak	ya
2	ya	ya	ya

Gambar 9.2 . Prototipe data training

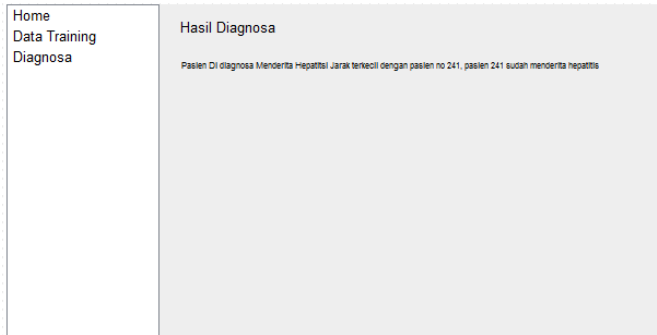
9.6.2 Tampilan Data Latih (mendiagnosa)



Data Training	
Muntah Minimal 3 kali	<input type="radio"/> Ya <input type="radio"/> Tidak
Demam 3 Hari	<input type="radio"/> Ya <input type="radio"/> Tiadak
Warna Mata Kuning	<input type="radio"/> Ya <input type="radio"/> Tiadak
BAK Kuning Teh	<input type="radio"/> Ya <input type="radio"/> Tiadak
Nafsu Makan Menurun	<input type="radio"/> Ya <input type="radio"/> Tiadak
Nyeri Perut Kanan Atas	<input type="radio"/> Ya <input type="radio"/> Tiadak
<input type="button" value="Diagnosa"/>	

Gambar 9.3. Prototipe data Diagnosa

9.6.3 Hasil Diagnosa



Gambar 9.4. Prototipe Hasil Diagnosa

DAFTAR PUSTAKA

- Agung, M, T. 2009 Penerapan Data Mining Pada Data Transaksi Penjualan Untuk Mengatur Penempatan Barang Menggunakan Algoritma Apriori,2009. Konsep Data Mining-Klasifikasi Pohon Kpeutusan, Gunadarma. Udinus
- Amir Amri, *Bunga Rampai Hukum Kesehatan*, Widya Medika, Jakarta, 1997.
- Bertalya,2009. Konsep Data Mining-Klasifikasi Pohon Kpeutusan, Gunadarma
- Fathasyah,2012, Basis Data, Bandung
- Hardjoeno UL. 2007. Kapita selekta hepatitis virus dan interpretasi hasil laboratorium. Makassar: Cahya Dinan Rucitra: hlm. 5-14
- Hartanto, Junaidi,2011 Data Transformation Pada Data Mining, Sekolah Tinggi Surabaya
- Hermawati, A, Fajar. 2013. Data Mining, Yogyakarta
- Infodatin, Situasi dan analisis Hepatitis, Pusat data dan informasi Kemenerian Kesehatan RI
- Kementrian Kesehatan Republik Indonesia. 2013. Laporan hasil riset kesehatan dasar Indonesia (Riskesdas). Jakarta: Badan Litbangkes. hlm.109-110

Lestari, Tri, 2009, Analisis Keranjang Belanja Pada Data Transaksi Penjualan, IPB

Maju Mandar, 2001 Hukum Kedokteran, bandung

Mustafa S, Kurniawaty E. 2013. Manajemen gangguan saluran cerna panduan bagi dokter umum. Lampung: Anugrah Utama Raharja(Aura).

Rham ,2011 Data Cleaning: Problems and Current Approaches, University of Leipzig, Germany

Thedja MD. 2012. Genetic diversity of hepatitis B virus in Indonesia: Epidemiological and clinical significance. Jakarta: DIC creative

Widiana M. E, 2010, Dasar-dasar Pemasaran Bndung

Wikepedia, <https://id.wikipedia.org/wiki/Variabel> diakses pada 29-09-2017

K-Nearest Neighbor (KNN)