PERTEMUAN 8

KNOWLEDGE ENGINEER

A. Tujuan Pembelajaran

Setelah mempelajari materi Sistem Penunjang Keputusan tentang Metode Akuisisi Knowledge, pada pertemuan ini, mahasiswa mampu menjelaskan tentang terjemahan informasi dan penyampaian kepada programmer yang membangun aplikasi.

B. Uraian Materi

1. Knowledge Engineer

Dengan meningkatnya kepedulian untuk pembangunan masa depan yang berkelanjutan, individu dan organisasi datang dengan pendekatan yang inovatif dan kreatif. Kecerdasan buatan telah menjadi pemain kunci baru-baru ini dalam mewujudkan dan mencapai tujuan masa depan yang berkelanjutan. Salah satu kasus penggunaan tersebut dapat ditemukan di "Masyarakat buatan untuk pengembangan sistem metropolitan yang terintegrasi dan berkelanjutan" yang telah mencoba untuk memecahkan masalah lalu lintas modern dengan menggunakan infrastruktur cerdas.

Pembelajaran Mesin adalah bidang ilmu komputer dan bagian dari kecerdasan buatan di mana komputer memperoleh potensi untuk belajar dan meningkatkan dari pengalaman tanpa diprogram secara eksplisit. Pembelajaran Mesin menggunakan model statistik untuk membuat prediksi berdasarkan data yang diumpankan sebelumnya. Rekayasa Data dan rekayasa pengetahuan memainkan peran kunci dalam mengimplementasikan analisis prediktif berdasarkan pembelajaran mesin. Pengetahuan adalah penipu

Cept yang berarti akumulasi informasi masuk akal yang dikumpulkan selama periode waktu tertentu. Salah satu keuntungan utama dari rekayasa pengetahuan adalah bahwa ia dapat dibagi di antara jaringan yang serupa untuk memperluas jangkauan jaringan tunggal ke kelompok yang lebih besar melalui proses yang disebut, kualitas perolehan pengetahuan dari jaringan terkait Kualitas akuisisi pengetahuan ini telah terbukti sangat ekonomis.

Referensi menyimpulkan sejumlah teknik pra-pemrosesan data untuk memecahkan masalah data , memahami sifat data, melakukan analisis data yang mendalam. Metode pra-pemrosesan data dibagi menjadi beberapa teknik berikut yaitu Pembersihan Data, Transformasi Data, dan Reduksi Data. Oleh karena itu, ini merupakan masalah penting sementara pra-pemrosesan dipertimbangkan untuk analisis data.

Masalah terpenting dalam pra-pemrosesan data adalah mengetahui informasi berharga apa yang ada dalam data mentah dan di mana menyimpannya. Keterlibatan pakar domain menghasilkan beberapa umpan balik yang berguna untuk verifikasi dan validasi dalam teknik pra-pemrosesan data tertentu. Umpan balik diperlukan dari proses analisis data utama. Proses yang lengkap dikaitkan dengan pekerjaan yang baik jika data yang diolah memberikan informasi yang berharga di kemudian hari proses analisis data.

Ladha dkk. telah menyebutkan keuntungan dari pemilihan fitur berikut:

- a. Kurangi dimensi untuk mengurangi kebutuhan penyimpanan dan meningkatkan kecepatan algoritme.
- b. Hapus data yang berlebihan, tidak relevan, atau berisik.
- c. Meningkatkan kualitas data.
- d. Meningkatkan akurasi model yang dihasilkan.
- e. Reduksi set fitur.
- f. Meningkatkan kinerja untuk mendapatkan akurasi prediksi.
- g. Dapatkan pengetahuan tentang prosesnya.

Tujuan pemilihan fitur dan ekstraksi fitur berkaitan dengan pengurangan dimensi untuk meningkatkan analisis data. Aspek penting menjadi relevan dalam mempertimbangkan dataset dunia nyata yang memiliki ratusan atau ribuan fitur. Penulis menyimpulkan untuk terlebih dahulu melakukan reduksi tanpa mengubahnya sedangkan ekstraksi fitur mengurangi dimensionalitas. Itu ditekankan untuk menciptakan fitur lain yang lebih signifikan dengan komputasi transformasi fitur unik .

Convolutional neural network (CNN) adalah jenis jaringan saraf pembelajaran dalam (NN) untuk pengenalan dan klasifikasi gambar. Tautan memiliki bobot bernilai yang disetel dengan baik dalam proses pelatihan yang menghasilkan jaringan yang terlatih. Jaringan saraf memiliki lapisan yang dihubungkan oleh neuron buatan. Setiap lapisan mengkategorikan satu set pola

input sederhana. Sebuah CNN standar memiliki 5-25 lapisan yang diakhiri dengan lapisan keluaran. Database besar gambar berkualitas baik dengan fitur yang kuat dan khas diambil untuk hasil klasifikasi yang baik. Model CNN pada akurasi klasifikasi menghilangkan kebutuhan Graphics Processing Unit (GPU) untuk pelatihan meskipun keuntungan mempersingkat waktu pelatihan dan pembelajaran transfer melebihi pelatihan penuh.

Deteksi malware berbasis perilaku otomatis menggunakan teknik pembelajaran mesin adalah metode yang bijaksana dan bermanfaat. Ini telah menghasilkan laporan perilaku

lingkungan yang diemulasi (kotak pasir) dan malware akan diidentifikasi secara otomatis. Pengklasifikasi yang berbeda yaitu k-Nearest Neighbours (kNN), Naïve Bayes, J48 Decision Tree, Support Vector Machine (SVM), dan Multilayer Perceptron Neural Network (MLP) digunakan. Hasil eksperimen yang dicapai oleh pohon keputusan J48 dengan recall 95.9%, false positive rate 2.4%, presisi 97.3%, dan akurasi 96.8% menyimpulkan bahwa penggunaan teknik pembelajaran mesin mendeteksi malware dengan cukup efisien dan efektif.

Para peneliti telah mengusulkan metodologi berbasis data yang terdiri dari metode CI untuk membandingkan dan meramalkan parameter AQ. Metode PCA dan ANN dipilih untuk tugas CI. Penerapan metode hybrid baru untuk memilih variabel input dari model JST-MLP dipertimbangkan. Untuk meningkatkan akurasi model peramalan, algoritma nonlinier seperti JST-MLP, pelatihan multi-lipat dan skema validasi diadopsi untuk polusi udara.

2. Tujuan Knowledge Engineer

Tujuan pemilihan fitur dan ekstraksi fitur berkaitan dengan pengurangan dimensi untuk meningkatkan analisis data. Aspek penting menjadi relevan dalam mempertimbangkan dataset dunia nyata yang memiliki ratusan atau ribuan fitur. Penulis menyimpulkan untuk terlebih dahulu melakukan reduksi tanpa mengubahnya sedangkan ekstraksi fitur mengurangi dimensionalitas. Itu ditekankan untuk menciptakan fitur lain yang lebih signifikan dengan komputasi transformasi fitur unik.

3. Proses Knowledge Engineer

Teknik pembelajaran yang diawasi telah digunakan untuk mengusulkan metode yang dapat membantu kita memahami indeks kualitas udara di berbagai negara bagian di India dan membuat model untuk memprediksi kualitas udara berdasarkan keberadaan Nitrogen Oksida (NO2) dan Materi Partikulat (PM10) atas dasar pemahaman dan pengetahuan. Ini juga dapat membantu kami untuk mengurangi area yang diminati yang memerlukan perhatian segera dalam hal kualitas udara.

Selain itu, model ini dapat diimplementasikan untuk mencapai tujuan yang lebih besar dari prediksi kualitas udara di seluruh dunia. Kami telah menggunakan pembelajaran yang diawasi untuk membuat model prediksi indeks kualitas udara suatu daerah dengan mengumpulkan informasi statistik dasar. Melalui metodologi pembelajaran yang diawasi serupa dalam klasifikasi citra pembelajaran yang dalam, metode lain telah diusulkan untuk pengelolaan limbah yang efektif dengan memperluas ruang lingkup pembelajaran yang dalam ke bidang lain seperti Internet of things [19-21]. Perkiraan kesalahan yang digunakan dalam makalah ini untuk memprediksi keakuratan model pembelajaran mesin adalah, Mean Absolute Error dan R Squared error

4. Contoh kasus Knowledge

Dalam kasus berikut, kumpulan data pemerintah telah dikumpulkan dari situs sumber pemerintah berkenaan dengan kandungan Nitrogen Dioksida (NO2) dan kandungan materi partikulat (PM10) yang ada di udara di berbagai negara bagian India untuk tahun 2011. Kehadiran konten ini berperan sangat penting, peran penting dalam menentukan kualitas udara

Langkah-langkah yang Terlibat:

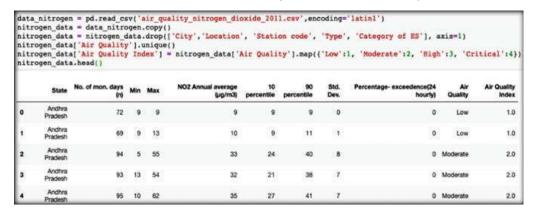
- a. Mengumpulkan Data Mentah
- b. Mengekstrak Data yang Diinginkan
- c. Data Pra-proses
- d. Visualisasikan Data
- e. Membandingkan Data dan Poin Pengetahuan
- f. Mengekstrak Informasi yang Benar
- g. Membuat model untuk analisis prediktif menggunakan pembelajaran mesin
- h. Validasi silang model dengan data masa depan

5. Proses Mengelola Pengetahuan

Data dikumpulkan melalui pengikisan situs web dan organisasi data terbuka seperti data.gov.in. Sumber URL yang disebutkan adalah situs web pemerintah otentik yang menyediakan data sah yang dikumpulkan oleh pemerintah India tentang berbagai topik seperti kualitas udara, kualitas air, dll

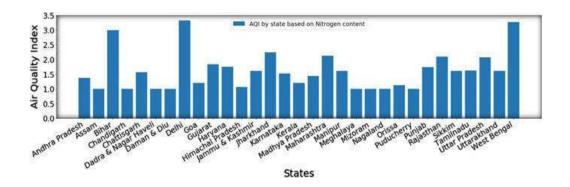
Ekstraksi data adalah proses membaca data berharga dari file data mentah termasuk csv, sql dll. Analisis hasil dilakukan dengan menggunakan grafik program yang dihasilkan di perpustakaan Bohek dan Matplotlib di python. Gambar 11 merepresentasikan data awal sebelum melakukan pembersihan data dan proses rekayasa data. Pada Gambar 11, kami telah mengambil kolom yang sesuai dengan data mentahnya yang paling diinginkan untuk melakukan proses pemilihan fitur sesuai kebutuhan kami

Data perlu diproses terlebih dahulu agar sesuai untuk tujuan visualisasi.



Gambar 10. Validasi Data

Data perlu dipahami sebelum mengambil poin pengetahuan valid yang diperlukan untuk melakukan rekayasa pengetahuan. Visualisasi data membuat proses rekayasa data dan ekstraksi pengetahuan menjadi proses yang nyaman. Data yang ditunjukkan grafik di bawah ini menunjukkan kualitas udara terhadap setiap negara bagian India pada skala 1-4 di mana 1 adalah kategori kondisi baik sedangkan 4 adalah kondisi sangat kritis



Gambar 11. AQI kandungan Nitrogen Dioksida (NO2) pada skala kritis 1-4.

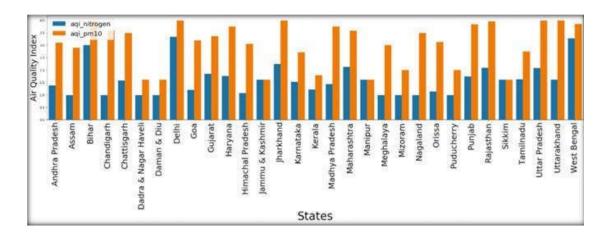
Grafik batang yang ditunjukkan pada Gambar 4, secara jelas mewakili tingkat kualitas udara di berbagai negara bagian. Visualisasi ini sangat membantu dalam memilih area yang diminati yang harus waspada untuk bekerja pada pengendalian kualitas udara.

Jelas terlihat bahwa negara bagian yang sangat tercemar termasuk Delhi, Benggala Barat dan Bihar yang memiliki kondisi udara paling kritis berdasarkan nitrogen dioksida dan materi partikulat. Informasi ini dapat digunakan oleh badan pengawas kualitas udara utama untuk mengurangi upaya kerja mereka di daerah yang sangat tercemar.

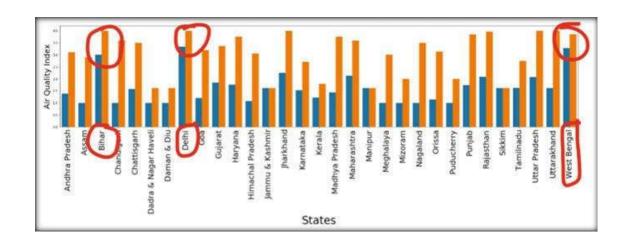
ditunjukkan bahwa dengan visualisasi data, dimungkinkan untuk mengidentifikasi negara bagian India yang memiliki kandungan nitrogen dioksida dan materi partikulat yang sangat tinggi.

Dalam langkah-langkah berikutnya, model prediksi telah dikembangkan untuk memprediksi kualitas udara berdasarkan informasi mengenai kandungan berbahaya tertentu. Proses visualisasi data ini dapat digunakan untuk mengatur.

Model pembelajaran mesin harus dilatih dengan data yang benar untuk menghindari kasus kekurangan atau kelebihan data. Karena data yang diperoleh dari.



Gambar 12. Negara dengan indeks kualitas udara PM10 dan NO2 yang rendah



	State	City	Location	Station Code	Туре	Category of ES	No. of mon. days (n)	Min	Max	PM10 Annual average (µg/m3)	10 percentile	90 percentile	Standard Deviation	Percentage- exceedence(24 hourly)	Air Quality
0	Andhra Pradesh	Chitoor	GNC Toll Gate Tirumala	582	RIRUO	NaN	103	24	52	40	35	45	4	0	Moderate
1	Andhra Pradesh	Guntur	Near Hindu College, Market Road	583	RIRuO	NaN	105	71	79	75	2	73	77	0	High
2	Andhra Pradesh	Hydrabad	Tarnaka, NEERI Lab. IICT Campus	150	RIRuO	NaN	88	8	70	26	17	35	10	0	Low
3	Andhra Pradesh	Hydrabad	Nacharam, Industrial Estate	151	RIRUO	NaN	87	12	100	30	19	42	15	0	Low

Tabel 9 Panduan untuk memilih sistem inferensi

Sumber daya pemerintah itu sendiri sangat bersih. Jadi untuk mencapai kesesuaian yang akurat dari konsep model pengurangan dimensi diterapkan hanya dengan menghapus data yang tidak relevan.

menunjukkan representasi tanpa melakukan korelasi apapun. Dapat disimpulkan bahwa Kode Negara, Kota, Lokasi dan Stasiun hanya akan mengakibatkan over fitting data karena tidak memiliki hubungan ilmiah dengan komposisi kimia komponen udara.

merupakan daftar fitur akhir yang telah dipilih berdasarkan pemilihan fitur statistik dan korelasi. Selanjutnya, fitur ini awalnya dinamai sebagai "Kualitas Udara" dan telah diubah namanya menjadi "Indeks Kualitas Udara". Tipe data fitur ini diubah dari integer menjadi string untuk klasifikasi yang lebih baik dan konten yang lebih relevan dalam hal tipe data intra

Setelah pemilihan fitur, studi data menyeluruh untuk model prediktif dirancang yang dapat mengklasifikasikan kualitas udara suatu daerah dalam hal keberadaan kandungan materi partikulat di daerah tertentu. Pendekatan klasifikasi regresi campuran digunakan. Ada empat label klasifikasi atas dasar mutu antara lain:

- a. Baik 1
- b. Sedang 2
- c. Tinggi 3
- d. Kritis 4.

dimana 1 menunjukkan kualitas udara terbaik

Pendekatan Model yang Digunakan:

a. Semua kolom berbasis lokasi dan kolom string ketat dijatuhkan.

No. of mon. days (n) PM10 Annual average (µg/m3) 10 percentile 90 percentile Standard Deviation Percentage- exceedence(24 hourly) Air Quality Inde

Gbr. 7 Kumpulan fitur akhir setelah melakukan korelasi fitur dan pemilihan fitur

- b. Mengisi kolom kosong indeks kualitas udara dengan rata-rata dari semua nilai kolom yang tersedia pada negara bagian tersebut.
- c. Dalam hal semua nilai kolom suatu negara kosong, maka nilai kolom tersebut diisi dengan rata-rata semua kolom nilai indeks kualitas udara.

d. Semua nilai kolom lainnya diisi dengan 0 default karena data yang dikumpulkan sudah berisi semua data statistik yang diperlukan untuk prediksi.

- e. Data kualitas udara berdasarkan materi partikulat dibagi menjadi set pengujian dan pelatihan. Data yang digunakan untuk tujuan ini adalah untuk tahun 2011.
- f. Model dilatih pada regresi atau algoritma menggunakan algoritma Regresi Hutan Acak pada data pelatihan untuk tahun 2011.
- g. Validasi k-fold cross digunakan untuk memverifikasi model pada split yang berbeda.
- h. Akhirnya model diuji pada data uji untuk tahun 2011. Model tersebut mampu mencapai akurasi sekitar 84%. seperti yang ditunjukkan pada Gambar 8.
- i. Kemudian dibuat prediksi regresi pada data tahun 2012 yang pertama kali diperkenalkan ke model. Dalam pendekatan ini, R² square 0,88 dicapai dan Mean Absolute Score 0,10 dicapai seperti yang ditunjukkan pada Gambar. 9.
- j. Untuk menyempurnakan hasil ini lebih lanjut, pendekatan klasifikasi diambil di mana keluaran yang diprediksi dibulatkan untuk mengurangi jarak antara prediksi aktual dan prediksi yang diusulkan. Setelah algoritma ini melakukan penyetelan skor R² dan Mean Absolute Error ditingkatkan seperti yang ditunjukkan pada Gambar 1

```
predict = regr.predict(X_test)
regr.score(X_test, y_test)
0.8440907077825055
```

6. Evaluasi Knowledge Engineering

Untuk mengevaluasi akurasi menggunakan mean absolute error, R^2 dan Mean Absolute Score telah ditentukan. Angka-angka yang disebutkan selanjutnya diperpanjang untuk membuat persentase akurasi yang dapat dilaporkan menggunakan akurasi kesalahan absolut rata-rata. Akurasi ini diperoleh dengan mendapatkan P

Perbedaan absolut antara nilai prediksi dan nilai aktual. Daftar nilai absolut ini selanjutnya digunakan untuk menemukan persentase akurasi model yang diusulkan Telah diamati bahwa pendekatan yang diusulkan mampu

mencapai akurasi 96% kira-kira pada kumpulan data baru yang tidak pernah diperkenalkan ke model saat melatihnya.

Keuntungan mencapai tujuan pembangunan berkelanjutan:

- a. Rekayasa Data dan Rekayasa Pengetahuan yang diterapkan melalui contoh di atas adalah alat yang sangat berguna untuk menemukan area yang menjadi perhatian yang membutuhkan perhatian segera untuk mencapai tujuan pembangunan berkelanjutan.
- b. Analisis prediktif dapat bekerja secara independen dari penginderaan perangkat elektronik untuk memprediksi masa depan tren perubahan iklim dengan mempelajari tren sebelumnya.
- c. Metode yang sangat hemat biaya dan membutuhkan intervensi manusia yang sangat sedikit setelah dilatih dengan data yang cukup.
- d. Dapat dengan mudah diperluas ke beberapa perangkat secara sinkron untuk mengumpulkan informasi.

Ini hanyalah salah satu dari banyak kasus penggunaan di mana pembelajaran mesin, bagian dari kecerdasan buatan, dapat digunakan untuk memahami bidang yang menjadi perhatian yang menjadi penghalang untuk mencapai tujuan pembangunan berkelanjutan

IoT dan Pembelajaran Mendalam untuk Pengawasan

Pembelajaran mendalam menyediakan algoritma yang sangat kuat untuk identifikasi dan klasifikasi gambar yang dapat dikombinasikan dengan teknologi tertentu lainnya untuk melaksanakan dan mencapai tujuan pembangunan berkelanjutan pada skala kecil, menengah dan besar dengan sangat efisien

Jaringan Neural Konvolusional (CNN)

CNN adalah metodologi berbasis model pembelajaran mendalam yang terutama digunakan untuk mengklasifikasikan gambar dan objek melalui cluster serupa. Algoritme jaringan saraf konvolusional pembelajaran mendalam bekerja pada konsep belajar dan mengklasifikasikan berdasarkan pengalaman. Terlepas dari klasifikasi gambar, jaringan saraf konvolusional memainkan peran penting dalam klasifikasi teks dan analitik teks.

CNN biasanya terdiri dari lapisan dasar berikut untuk menarik kesimpulan untuk analisis teks:

- a. Lapisan Masukan
- b. Lapisan Tersembunyi
- c. Pooling Layer
- d. Max Pooling Layer
- e. Konv + RELU
- f. Lapisan Keluaran (Lapisan Klasifikasi)
- g. Ratakan Lapisan
- h. Lapisan Terhubung Sepenuhnya
- i. Lapisan Soft Max. Internet of Things (IoT)

Internet of things adalah bidang teknik dan ilmu komputer di mana banyak hal saling berhubungan melalui WLAN internet. Dalam pengertian yang paling sederhana, IoT adalah seperangkat perangkat elektronik yang terus berkomunikasi satu sama lain untuk menjalankan tugas tertentu untuk mencapai hasil akhir.

Metode yang Diusulkan untuk Mencapai Tujuan Pembangunan Berkelanjutan:

- a. Air Bersih dan Sanitasi
- b. Kesehatan dan Kesejahteraan yang Baik
- c. Kehidupan Di Bawah Air
- d. Kehidupan di Darat
- e. Komponen Arsitektur
- f. Pembelajaran Mendalam untuk Klasifikasi Gambar
- g. Drone untuk pengawasan dan observasi jarak jauh
- h. Alat analisis untuk melatih lebih banyak data.

Langkah-langkah yang Terlibat:

- a. Latih model CNN offline untuk klasifikasi gambar dari berbagai jenis limbah material.
- b. Menulis program untuk mengontrol drone menggunakan laptop dan menerima gambar secara real time yang diambil melalui kamera drone.
- c. Interfacing drone untuk menyimpan data citra.
- d. Mengumpankan data gambar ke pengklasifikasi CNN untuk mendapatkan pembaruan rutin dari jenis sampah di sekitar suatu area.

Keuntungan:

a. Arsitektur yang diusulkan bisa sangat efisien dalam mensurvei jenis-jenis sampah di sekitar area yang tidak dapat diakses manusia. Karena ukuran drone yang kompak dan kemampuan untuk mengontrol jangkauan jarak jauh menjadi tugas yang lebih mudah.

- b. Dengan membuat kapal selam kecil yang setara dengan drone, pengelolaan limbah bawah air dapat mengurangi jumlah uang yang dikeluarkan untuk pengawasan limbah bawah air.
- c. Metode yang diusulkan dapat secara eksponensial mengurangi waktu dan biaya tenaga kerja manusia dan dapat dengan cepat digunakan untuk menentukan bidang yang diminati untuk pengelolaan limbah yang efektif.

Memperluas cakupan jaringan saraf untuk pendidikan sampah lokal:

Model jaringan saraf yang dilatih dalam metode yang diusulkan di atas dapat diperluas untuk membuat jaringan terpusat yang lebih besar untuk pemantauan dan pembuangan sampah dan sampah untuk area yang lebih bersih.

C. Soal Latihan/Tugas

- 1. Syarat yang harus dimiliki seorang knowledge engineer?
- 2. Jelaskan teknik-teknik representasi pengetahuan yang biasa digunakan dalam pengembangan suatu sistem pakar?

D. Referensi

- Chichilnisky G (2000) An axiomatic approach to choice under uncertainty with catastrophic risks. Resour Energy Econ 22(3):221–231
- Lele SM (1991) Sustainable development: a critical review. World Dev 19(6):607-621
- Robert KW, Thomas MP, Leiserowitz AA (2005) What is sustainable development? Goals, indicators, values, and practice. Environ: Sci Policy Sustain Dev 47(3):8–21
- Yun, JH et al (2016) Not deep learning but autonomous learning of open innovation for sus- tainable artificial intelligence. Sustainability 8(8):797
- Dean J, Corrado G, Monga R, Chen K, Devin M, Mao M, Ng AY (2012) Large scale

distributed deep networks. In: Advances in neural information processing systems, pp 1223–1231

- Spring R, Shrivastava A (2017) Scalable and sustainable deep learning via randomized hashing. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. ACM
- Michie D, Spiegelhalter DJ, Taylor CC (1994) Machine learning. Neural Stat Classif
- Studer R, Benjamins VR, Fensel D (1998) Knowledge engineering: principles and meth- ods. Data Knowl Eng 25(1–2):161–197

GLOSARIUM

Inferencing adalahlangkah-langkah dalam penalaran, bergerak dari premis ke konsekuensi logis; Secara etimologis, kata menyimpulkan berarti "meneruskan"

Inexact Reasoning adalah Penalaran Tak Pasti

Exact Reasoning adalah Penalaran Pasti.