

MAKALAH KELOMPOK 5

Fungsi Klasifikasi



Dosen Pengampu :
Widya Darwin S.Pd., M.Pd.T

Anggota Kelompok 5 :

1. Jimmy Mahendra (22346009)
2. Najwa Alawiyah Siregar (22346040)

PRODI INFORMATIKA
DEPERTEMEN ELEKTRONIKA
FAKULTAS TEKNIK
UNIVERSITAS NEGERI PADANG

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa, atas segala rahmat dan karunia-Nya sehingga kami dapat menyelesaikan makalah yang berjudul “Fungsi Klasifikasi” dengan baik dan tepat waktu. Makalah ini disusun guna memenuhi salah satu tugas mata kuliah Data Mining.

Dalam penyusunan makalah ini, kami telah berusaha menyajikan informasi yang jelas dan sesuai dengan topik yang dibahas, yaitu jenis-jenis penelitian berdasarkan fungsinya. Kami berharap makalah ini dapat memberikan wawasan serta pemahaman yang lebih mendalam mengenai Fungsi Klasifikasi dalam Data Mining dan Implementasinya dalam Berbagai Bidang.

Kami menyadari bahwa makalah ini masih memiliki kekurangan. Oleh karena itu, kami sangat mengharapkan kritik dan saran yang membangun dari pembaca untuk penyempurnaan di masa yang akan datang.

Kiranya makalah ini dapat bermanfaat bagi kita semua, terima kasih.

Painan, 26 Oktober 2024

Penulis,

Kelompok 5

DAFTAR ISI

KATA PENGANTAR.....	i
DAFTAR ISI.....	ii
BAB 1	1
PENDAHULUAN	1
1.1. LATAR BELAKANG	1
1.2. RUMUSAN MASALAH.....	2
1.3 TUJUAN	2
BAB II.....	3
PEMBAHASAN	3
2.1. PENJELASAN FUNGSI KLASIFIKASI	3
2.2. PERBANDINGAN ALGORITMA PENGKLASIFIKASI DATA	4
BAB III.....	8
KESIMPULAN	8
3. KESIMPULAN	8
DAFTAR PUSTAKA	9

BAB 1

PENDAHULUAN

1.1. LATAR BELAKANG

Data mining atau lebih di kenal juga dengan sebutan knowledge discovery in databases (KDD). Data mining merupakan salah satu cara yang digunakan untuk mendapatkan pengetahuan baru dengan memanfaatkan jumlah data yang sangat besar. Kegiatan dalam Data mining meliputi pengumpulan dan pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data yang berukuran besar. Beberapa teknik telah dikembangkan dan diimplementasikan untuk mengekstrak pengetahuan dan informasi untuk menemukan pola pengetahuan yang mungkin berguna untuk pengambilan keputusan. Teknik-teknik yang digunakan untuk pengekstrakan pengetahuan dalam data mining adalah pengenalan pola, clustering, asosiasi, prediksi dan klasifikasi.

Salah satu pendekatan yang dapat digunakan untuk menganalisis sekumpulan data adalah klasifikasi. Klasifikasi merupakan salah satu teknik data mining yang digunakan untuk membangun suatu model dari sampel data yang belum terklasifikasi untuk digunakan mengklasifikasi sampel data baru ke dalam kelas-kelas yang sejenis. Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu. Keluaran yang dihasilkan oleh klasifikasi data mining dapat digunakan untuk memperbaiki pengambilan keputusan bagi analis.

Klasifikasi termasuk ke dalam supervised learning karena menggunakan sekumpulan data untuk dianalisis terlebih dahulu, kemudian pola dari hasil analisis tersebut digunakan untuk pengklasifikasian data uji. Proses klasifikasi data terdiri dari pembelajaran dan klasifikasi. Pada pembelajaran data training dianalisis menggunakan algoritma klasifikasi, selanjutnya pada klasifikasi digunakan data testing untuk memastikan tingkat akurasi dari rule klasifikasi yang digunakan. Teknik klasifikasi dibagi menjadi lima kategori berdasarkan perbedaan konsep matematika, yaitu berbasis statistik, berbasis jarak, berbasis pohon keputusan, berbasis jaringan syaraf, dan berbasis rule. Ada banyak algoritma dari masing-masing kategori tersebut,

namun yang populer dan sering digunakan diantaranya yaitu naive bayes, nearest neighbour dan decision tree.

1.2. RUMUSAN MASALAH

Berdasarkan latar belakang di atas, masalah yang akan dibahas dalam makalah ini adalah:

1. Apa itu Fungsi Klasifikasi?
2. Apa saja perbandingan dari algoritma Pengklasifikasian naive bayes, nearest neighbour, dan decision tree?

1.3. TUJUAN

Adapun tujuan dari makalah ini adalah:

1. Memberikan pemahaman menyeluruh tentang fungsi dan peran teknik data mining, khususnya klasifikasi, dalam mengekstrak informasi berharga dari kumpulan data besar.
2. Menyampaikan informasi tentang keunggulan dan tantangan dari beberapa algoritma klasifikasi yang populer untuk memandu pilihan metode yang paling tepat sesuai kebutuhan aplikasi.
3. Menjelaskan metode dan algoritma yang sering digunakan dalam Fungsi Klasifikasi, serta penerapan masing-masing dalam berbagai konteks, seperti prediksi dan pengelompokan.

BAB II

PEMBAHASAN

2.1. PENJELASAN FUNGSI KLASIFIKASI

Pada persoalan klasifikasi, kita memiliki sejumlah kasus (sampel data) dan ingin memprediksi beberapa class yang ada pada sampel data tersebut. Tiap instan data berisi banyak atribut, dimana masing-masing atribut satu dari beberapa kemungkinan nilai. Hanya satu atribut diantara banyak atribut tersebut yang disebut dengan atribut target, sedangkan atribut yang lain disebut sebagai atribut predictor. Tiap kemungkinan nilai yang dimiliki oleh atribut target menunjukkan class yang diprediksi berdasarkan nilai-nilai dari atribut predictor. Metode klasifikasi digunakan untuk membantu dalam memahami pengelompokan data.

Klasifikasi merupakan salah satu metode dari data mining. Klasifikasi adalah metode prediktif yang melakukan pembelajaran terhadap data-data yang sudah ada sehingga menghasilkan suatu model yang digunakan untuk memprediksi data-data baru. Klasifikasi data terdiri dari dua proses yaitu tahap pembelajaran dan tahap pengklasifikasian. Tahap pembelajaran merupakan tahapan dalam pembentukan model klasifikasi, sedangkan tahap pengklasifikasian merupakan tahapan penggunaan model klasifikasi untuk memprediksi label kelas dari suatu data. Contoh sederhana dari teknik data mining klasifikasi adalah pengklasifikasian hewan berdasarkan atribut jumlah kaki, habitat dan organ pernafasannya akan diklasifikasikan ke dalam dua label kelas yaitu unggas dan ikan. Label kelas unggas adalah data yang memiliki jumlah kaki dua, habitatnya di darat, dan organ pernafasannya menggunakan paru-paru, sedangkan label kelas ikan adalah data yang memiliki jumlah kaki nol (tidak memiliki kaki), habitat di air, dan organ pernafasannya menggunakan insang.

2.2. PERBANDINGAN ALGORITMA PENGKLASIFIKASI DATA

Banyak algoritme yang dapat digunakan dalam pengklasifikasian data, namun dalam penelitian ini hanya akan membandingkan tiga algoritma saja, yakni naive bayes, nearest neighbour, dan decision tree.

A. K-Nearest Neighbor (KNN)

Nearest Neighbour adalah algoritma pengklasifikasian yang didasarkan pada analogi, yaitu membandingkan data uji dengan data pelatihan yang berada dekat dengan dan memiliki kemiripan dengan data uji tersebut. KNN merupakan metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikan berdasarkan mayoritas dari kategori pada KNN. Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan training sample. Algoritma KNN sangatlah sederhana, bekerja berdasarkan jarak terpendek dari query instance ke training sample untuk menentukan KNN-nya dan mudah untuk di implementasikan. KNN memiliki kemampuan kerja yang rendah ketika training dataset besar. Salah satu masalah pada algoritma ini adalah bobot yang sama dari semua atribut dalam menghitung jarak antara data testing dan data training, bagaimana pun, mungkin dari semua atribut ada beberapa atribut yang kurang penting untuk proses klasifikasi dan ada beberapa atribut yang lebih penting untuk proses klasifikasi. Sehingga tidak jelas jarak mana yang harus digunakan dan atribut mana yang harus digunakan untuk mendapatkan hasil terbaik. Hal ini dapat menyedatkan proses klasifikasi dan dapat menurunkan akurasi dari klasifikasi. Pendekatan yang banyak dilakukan untuk mengatasi masalah ini adalah dengan memberi bobot yang berbeda pada tiap-tiap atribut ketika mengukur jarak dua record. Pembobotan berguna untuk menentukan jarak antar atribut tetangga dengan record baru berdasarkan similarity.

Ketepatan algoritma k-NN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevan isinya terhadap klasifikasi. Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur, agar performa klasifikasi menjadi lebih baik. KNN juga merupakan contoh teknik lazy learning, yaitu teknik yang menunggu sampai pertanyaan (query) datang agar sama dengan data training

B. Naive Bayes

Naive Bayes merupakan pengklasifikasian dengan metode probabilitas yang ditemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes. Teorema tersebut dikombinasikan dengan naive di mana diasumsikan kondisi antar petunjuk (atribut) saling bebas. Klasifikasi naive Bayes diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Salah satu pengaplikasian dari naive Bayes yaitu pada bidang kesehatan.

Teorema bayes adalah perhitungan statistik dengan menghitung probabilitas kemiripan kasus lama yang ada dibasis kasus dengan kasus baru. Teorema bayes memiliki tingkat akurasi yang tinggi dan kecepatan yang baik ketika diterapkan pada database yang besar. Naïve bayes termasuk ke dalam pembelajaran supervised, sehingga pada tahapan pembelajaran dibutuhkan data awal berupa data pelatihan untuk dapat mengambil keputusan. Pada tahapan pengklasifikasian akan dihitung nilai probabilitas dari masing-masing label kelas yang ada terhadap masukan yang diberikan. Label kelas yang memiliki nilai probabilitas paling besar yang akan dijadikan label kelas data masukan tersebut. Naive bayes merupakan perhitungan teorema bayes yang paling sederhana, karena mampu mengurangi kompleksitas komputasi menjadi multiplikasi sederhana dari probabilitas. Selain itu, algoritma naive bayes juga mampu menangani set data yang memiliki banyak atribut. Model Naïve Bayes adalah klasifikasi statistik yang dapat digunakan untuk memprediksi suatu kelas. Model Naïve Bayes dapat diasumsikan bahwa efek dari suatu nilai atribut sebuah kelas yang diberikan adalah bebas dari atribut-atribut lain.

Kelebihan yang dimiliki oleh Naïve Bayes adalah dapat menangani data kuantitatif dan data diskrit, Naïve Bayes kokoh terhadap noise, Naïve Bayes hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter yang dibutuhkan untuk klasifikasi, Naïve Bayes dapat menangani nilai yang hilang dengan mengabaikan instansi selama perhitungan estimasi peluang, Naïve Bayes cepat dan efisiensi ruang.

C. Decision Tree

Algoritma decision tree merupakan algoritma yang umum digunakan untuk pengambilan keputusan. Decision tree akan mencari solusi permasalahan dengan menjadikan kriteria sebagai node yang saling berhubungan membentuk seperti struktur pohon. Decision tree adalah model prediksi terhadap suatu keputusan menggunakan struktur hirarki atau pohon. Setiap pohon memiliki cabang, cabang mewakili suatu atribut yang harus dipenuhi untuk menuju cabang selanjutnya hingga berakhir di daun (tidak ada cabang lagi). Konsep data dalam decision tree adalah data dinyatakan dalam bentuk tabel yang terdiri dari atribut dan record. Atribut digunakan sebagai parameter yang dibuat sebagai kriteria dalam pembuatan pohon

Decision tree adalah bentuk sederhana teknik klasifikasi pada sekumpulan kelas tak berhingga yang direpresentasikan ke dalam bentuk simpul (node) dan rusuk (edge). Biasanya, Decision Tree dipilih untuk menyelesaikan masalah dengan output yang bernilai diskrit.

Decision tree merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode decision tree mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk basis data seperti Structure Query Language (SQL) untuk mencari record pada data tertentu. Sebuah decision tree adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Pada decision tree setiap simpul daun menandai label kelas. Simpul yang bukan simpul akhir terdiri dari akar dan simpul internal yang terdiri dari kondisi tes atribut pada sebagian record yang mempunyai karakteristik yang berbeda. Simpul akar dan simpul internal ditandai dengan bentuk oval dan simpul daun ditandai dengan bentuk segi empat.

Ada beberapa konsep dalam decision tree, antara lain:

- a. Data dinyatakan dalam bentuk Tabel dengan atribut dan record.
- b. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan tree. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin dan temperatur. Salah satu atribut merupakan atribut yang menyatakan data solusi per-item data yang disebut dengan target atribut.
- c. Atribut memiliki nilai-nilai yang dinamakan dengan instance.

BAB III

KESIMPULAN

A. KESIMPULAN

Data mining merupakan teknologi yang sangat penting dalam dunia yang semakin bergantung pada data. Dengan berbagai fungsi utama seperti estimasi, prediksi, klasterisasi, klasifikasi, dan asosiasi, data mining memungkinkan pengambilan keputusan yang lebih baik melalui pemahaman pola-pola tersembunyi dalam data. Klasifikasi, sebagai salah satu teknik data mining yang paling umum, terbukti sangat berguna untuk mengkategorikan data baru berdasarkan model yang dilatih pada data historis. Proses ini melibatkan pembelajaran menggunakan data training untuk menciptakan model prediktif, yang kemudian diaplikasikan pada data baru guna menghasilkan prediksi akurat.

Metode klasifikasi seperti K-Nearest Neighbor (KNN), Naïve Bayes, dan Decision Tree menawarkan pendekatan yang berbeda dengan kelebihan dan kekurangannya masing-masing. KNN cocok untuk kasus sederhana tetapi dapat menjadi lambat dengan jumlah data yang besar, sedangkan Naïve Bayes sangat efisien dan memiliki kecepatan tinggi dalam menangani data besar. Sementara itu, Decision Tree sering dipilih untuk permasalahan keputusan kompleks karena kemampuannya membuat keputusan berdasarkan hirarki kriteria yang jelas.

Secara keseluruhan, klasifikasi dan fungsi lain dalam data mining mampu meningkatkan efisiensi, ketepatan, dan kecepatan dalam berbagai bidang, membantu organisasi untuk lebih mudah memperoleh wawasan dari data besar yang mereka miliki.

DAFTAR PUSTAKA

- Defiyanti, S., & Jajuli, M. (2015). *Integrasi Metode Klasifikasi Dan Clustering dalam Data Mining*. Konferensi Nasional Informatika (KNIF), 39-44.
- Dewi, S. (2016). *KOMPARASI 5 METODE ALGORITMA KLASIFIKASI DATA MINING PADA PREDIKSI KEBERHASILAN PEMASARAN PRODUK LAYANAN PERBANKAN*. *Jurnal Techno Nusa Mandiri*, 13(1), 60-65.
- Penerapan Algoritma Naive Bayes untuk Mengklasifikasikan Data Nasabah Asuransi*. Bustami. Aceh : *TECHSE Jurnal Penelitian Teknik Informatika*. 5.
- Sabransyah, M., Nasution, Y. N., & Amijaya, F. D. T. (2017). *Aplikasi Metode Naive Bayes dalam Prediksi Risiko Penyakit Jantung*. *JURNAL EKSPONENSIAL*, 8(2), 111-118.
- Shukla, A.Tiwari, R., & Kala, R. 2010. *RealLife Application of Soft Computing*. Taylor and Francis Groups, LLC.