

TUGAS METODE DATA MINING

Disusun untuk memenuhi salah satu tugas mata kuliah

Data Mining

Dosen : Muh. Hajar Akbar, ST., M.Kom



Disusun Oleh :

Nama : Miftahul Jannah

Nim : A1 20043

Prodi : Teknik Informatika

**UNIVERSITAS NAHDLATUL ULAMA
SULAWESI TENGGARA**

TUGAS!

1. Sebutkan dan jelaskan 5 teknik / metode data mining
2. Cari paper penelitian yang membahas tentang Teknik metode data mining sesuai dengan yang anda sebutkan
3. Review paper tersebut sesuai kaidah review paper

Penyelesaian!

1. Neural Network in Data Mining

Neural network dalam data mining adalah metode klasifikasi yang mengambil input, melatih dirinya sendiri untuk mengenali pola data input dan memprediksi output untuk input baru sejenis.

Jaringan saraf membentuk dasar pembelajaran mendalam, subbidang pembelajaran mesin yang berada di bawah kecerdasan buatan.

Perancangan algoritma jaringan saraf terinspirasi dari struktur otak manusia.

Sama seperti otak manusia yang bertanggung jawab atas kecerdasan dan daya pembedanya, jaringan saraf juga meniru otak manusia dan belajar dari pengalamannya serta menerapkan pembelajaran ini untuk klasifikasi dan prediksi.

Apa itu Jaringan Neural?

Jaringan saraf adalah sekumpulan unit input/output yang terhubung. Bobot dikaitkan dengan setiap koneksi, yang disesuaikan oleh jaringan selama proses pembelajaran. Sehingga, jaringan dapat memprediksi label kelas yang benar dari input tupel.

Jaringan saraf tiruan dirancang setelah struktur otak manusia. Struktur tersebut memiliki unit input/output yang menggambarkan neuron otak manusia yang kita sebut sebagai node dan kita mengacu pada koneksi

antar node sebagai penghubung. Gambar di bawah ini mewakili jaringan neuron buatan.

Bagaimana Jaringan Neural Bekerja?

Mengetahui struktur keseluruhan jaringan saraf memungkinkan kita memahami cara kerja jaringan saraf dengan bantuan contoh sederhana.

Apakah Anda tahu bagaimana kami dapat mengklasifikasikan bunga dan daun? Misalkan kita melihat sebuah gambar bunga yang tersusun dari 20 x 20 piksel yaitu sama dengan 400 ratus piksel

Node input jaringan menerima satu set nilai input. Jumlah node input sama dengan jumlah variabel input.

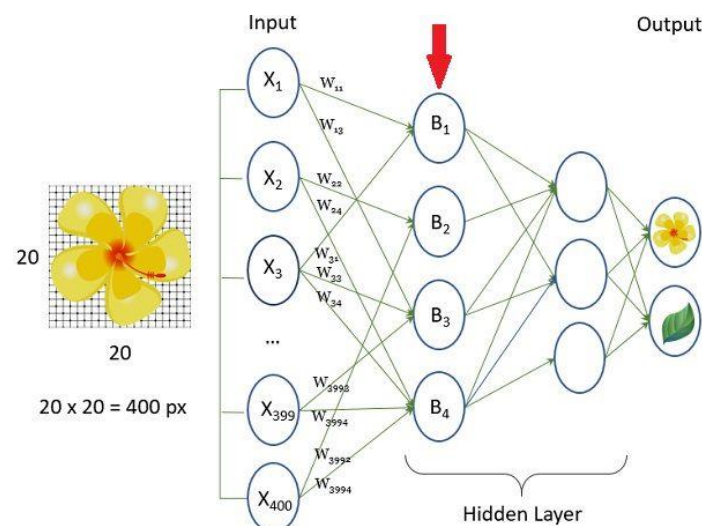
Sekarang masing-masing piksel ini dimasukkan sebagai masukan ke lapisan pertama dari jaringan saraf yaitu dalam contoh kita X 1 sampai X 400 .

Lapisan pertama jaringan terhubung ke lapisan berikutnya dengan tautan dan setiap tautan dikaitkan dengan bobot.

Sekarang, untuk mengevaluasi nilai setiap neuron di lapisan berikutnya, kami menerapkan fungsi penjumlahan.

Katakanlah untuk mengevaluasi nilai neuron pertama di lapisan kedua dari contoh kita, kita akan menerapkan fungsi penjumlahan:

$$= [X_1 \cdot P_{11} + X_3 \cdot P_{31}]$$



Hasil penjumlahan ini dikirim ke neuron pertama pada layer berikutnya. Sekarang setiap neuron di lapisan tersembunyi memiliki nilai numerik yang terkait dengannya yang kami sebut sebagai 'bias'.

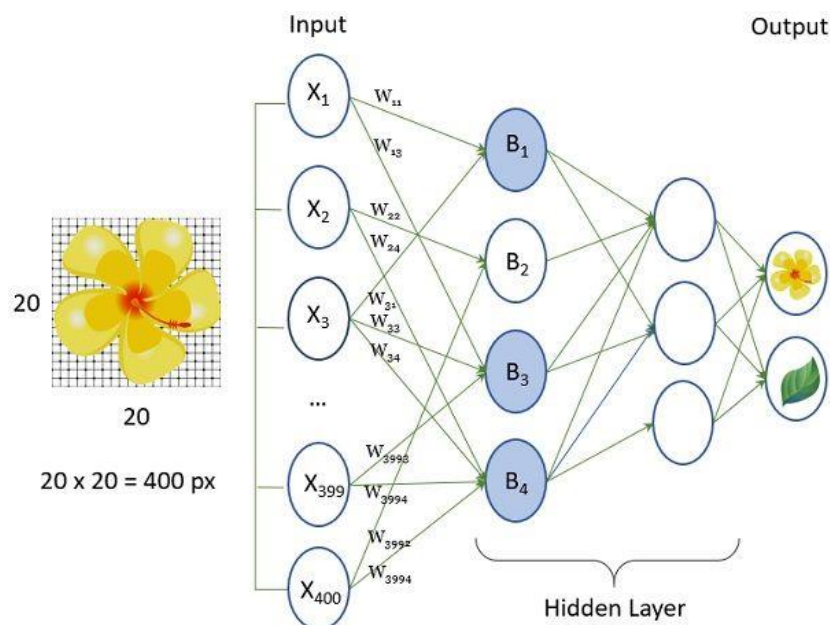
Jadi di sini, hasil penjumlahan ditambahkan ke bias B_1 .

$$= [X_1 * L_{11} + X_3 * L_{31}] + B_1$$

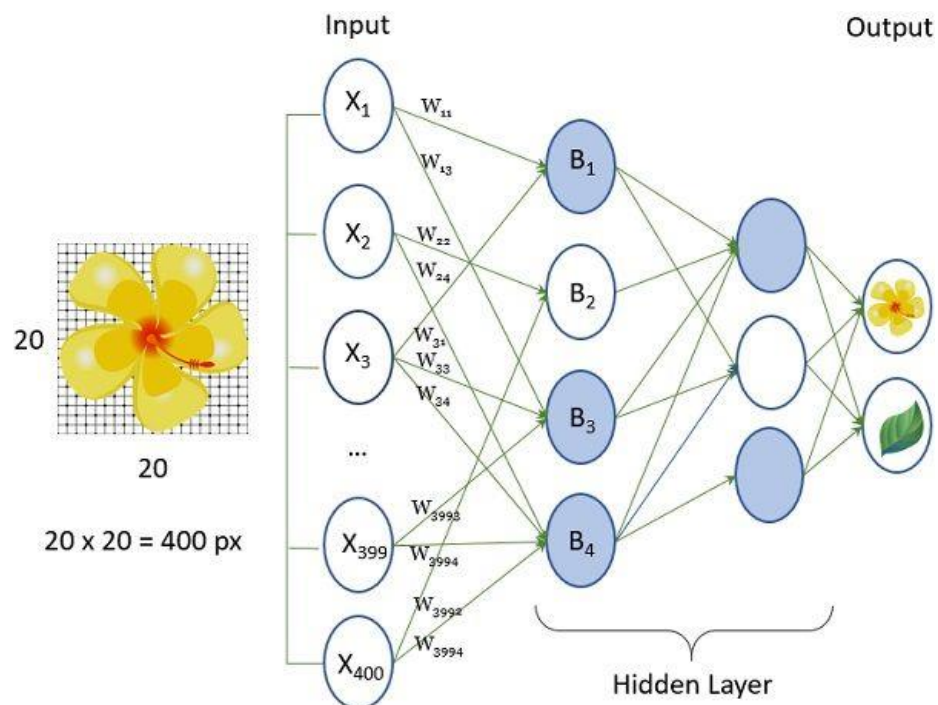
Hasil ini selanjutnya diteruskan ke fungsi aktivasi ambang yang memutuskan apakah neuron tertentu ini akan diaktifkan atau tidak.

$$\text{Keluaran} = \text{Fungsi Aktivasi} ([X_1 * W_{11} + X_3 * W_{31}] + B_1)$$

Neuron yang diaktifkan selanjutnya mengirimkan informasi ini ke neuron yang ada di lapisan berikutnya.

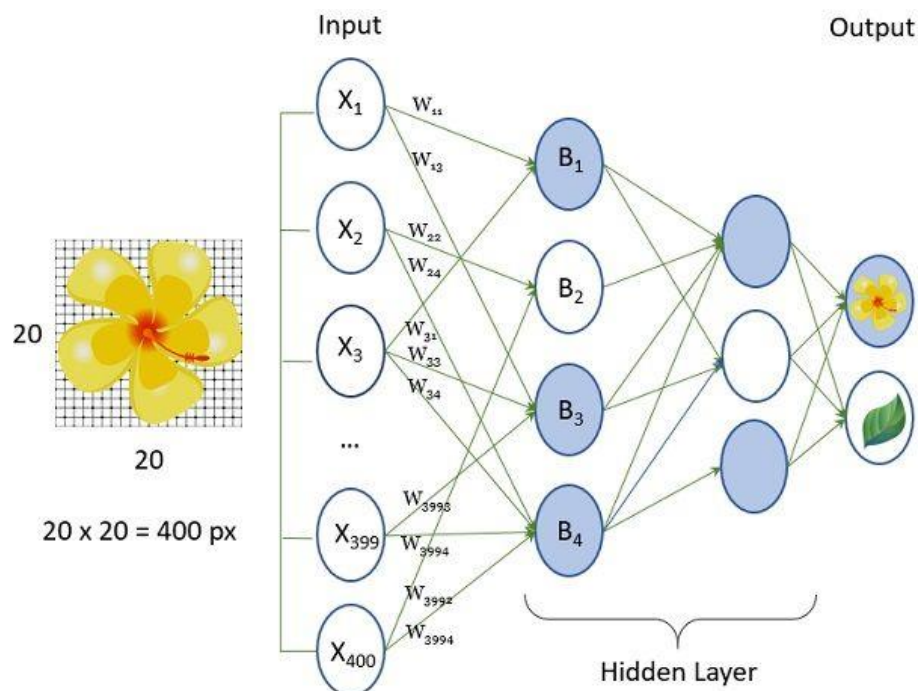


Dengan cara ini, informasi disebarkan melalui jaringan melalui tautan dan kami menyebutnya sebagai penyebaran penerusan.



Saat informasi mencapai lapisan keluaran, neuron dengan nilai tertinggi menentukan hasilnya. Nilai-nilai neuron di lapisan luar hanyalah probabilitas.

Neuron dengan probabilitas tertinggi pada lapisan keluaran adalah keluaran yang diprediksi oleh jaringan syaraf tiruan. Dalam kasus kami, neuron yang terkait dengan bunga memiliki probabilitas tertinggi untuk menggambarkan jawaban yang benar.



Kadang-kadang ketika jaringan belum dilatih, itu mungkin memprediksi jawaban yang salah. Nah untuk mengetahui hal ini, jaringan juga diumpankan bersama dengan outputnya. Jaringan membandingkan output yang diprediksi dengan output aktual. Untuk mengoreksi keluaran, bobot jaringan disesuaikan hingga jaringan memprediksi jawaban yang benar.

Penerapan Neural Network dalam Data Mining

Jaringan saraf dalam penambangan data melakukan tugas seperti:

- Klasifikasi
- Clustering atau kategorisasi
- Ramalan
- Perkiraan fungsi
- Optimasi
- Pengambilan oleh konten atau control

Jenis Jaringan Syaraf Tiruan

Ada beberapa jenis jaringan saraf tergantung pada perilakunya, yaitu:

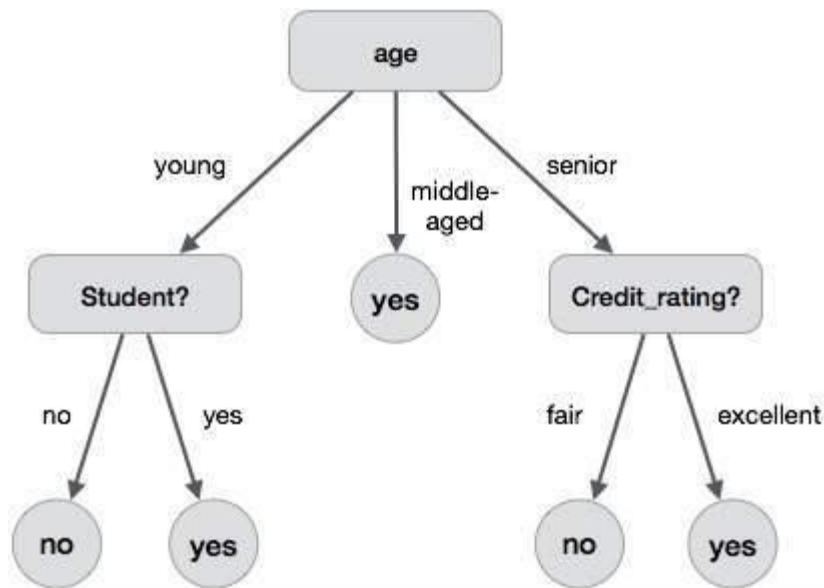
- Jaringan Syaraf Berulang
- Jaringan Syaraf Konvolusi
- Jaringan Syaraf Fungsi Basis Radial
- Jaringan Neural Umpan Maju
- Jaringan Saraf Modular

Decision Tree (DT)

Data Mining - Decision Tree Induction

Pohon keputusan adalah struktur yang mencakup simpul akar, cabang, dan simpul daun. Setiap simpul internal menunjukkan pengujian pada atribut, setiap cabang menunjukkan hasil pengujian, dan setiap simpul daun memegang label kelas. Simpul paling atas di pohon adalah simpul akar.

Pohon keputusan berikut adalah untuk konsep buy_computer yang menunjukkan apakah seorang pelanggan di sebuah perusahaan kemungkinan besar akan membeli komputer atau tidak. Setiap node internal mewakili tes pada atribut. Setiap simpul daun mewakili kelas.



Manfaat memiliki pohon keputusan adalah sebagai berikut :

- Itu tidak memerlukan pengetahuan domain apa pun.
- Mudah untuk dipahami.
- Langkah-langkah pembelajaran dan klasifikasi pohon keputusan sederhana dan cepat.

Decision Tree Induction Algorithm

Seorang peneliti mesin bernama J. Ross Quinlan pada tahun 1980 mengembangkan algoritma pohon keputusan yang dikenal sebagai ID3 (Iterative Dichotomiser). Kemudian, ia mempresentasikan C4.5, yang merupakan penerus dari ID3. ID3 dan C4.5 mengadopsi pendekatan serakah. Dalam algoritma ini, tidak ada backtracking; pohon-pohon dibangun dengan cara membagi-dan-menaklukkan rekursif dari atas ke bawah.

Rule Induction (RI)

DATA MINING WITH RULE INDUCTION

kita telah menyelidiki cara untuk mempersiapkan data hukum untuk fase data mining. Tugas-tugas sebelum data mining termasuk:

1. Seleksi dan pra-pemrosesan data – ini mengacu pada pra-pemrosesan data untuk menghilangkan atau mengurangi noise dan penanganan nilai yang hilang.
2. Transformasi data – mengubah data-set agar fase data mining dapat lebih efektif.

Dalam bab ini, kami terutama fokus pada satu pendekatan klasifikasi dan prediksi data mining, yaitu induksi aturan. Pada bab berikutnya, kami membahas aturan asosiasi, logika fuzzy, dan klasifikasi Bayesian. Teknik data mining lainnya, yaitu jaringan saraf, dibahas dalam bab tujuh.

Data mining dapat dilihat sebagai mekanisme otomatis untuk mempelajari pola dari dataset. [Michalski dan Larson 1978] mengidentifikasi strategi pembelajaran umum berikut:

- a) Implantasi Langsung Pengetahuan,
- b) Pembelajaran dari Instruksi,
- c) Pembelajaran dengan Inferensi Deduktif,
- d) Pembelajaran dengan Analogi,
- e) Pembelajaran dari Contoh,
- f) Pembelajaran dengan Pengamatan dan Penemuan.

Strategi pembelajaran a-f disajikan berdasarkan urutan peningkatan usaha dari pembelajar dan penurunan jumlah usaha dari pengajar.

Urutan ini juga mencerminkan peningkatan kesulitan dalam membangun

sistem pembelajaran yang mampu menggunakan strategi pembelajaran yang diberikan.

Semakin tinggi strategi pembelajaran, semakin banyak inferensi, dan oleh karena itu semakin banyak biaya dan usaha (oleh pembelajar) yang terlibat dalam menurunkan pengetahuan yang diinginkan. Pemerolehan pengetahuan mesin menjadi lebih mudah jika kita tahu cara menyelesaikan masalah dengan tepat. Dalam hal ini, mengajarkan komputer melalui instruksi lebih sederhana dan lebih baik daripada melibatkannya dalam proses pembelajaran induktif. Namun, ada banyak domain aplikasi di mana algoritma yang tepat tidak diketahui atau sulit diperoleh. Hukum adalah salah satu domain tersebut.

Pembelajaran induktif adalah proses memperoleh pengetahuan dengan mengambil inferensi induktif dari fakta-fakta yang diberikan oleh pengajar atau lingkungan. Proses ini melibatkan operasi generalisasi, transformasi, koreksi, dan penyempurnaan pengetahuan.

k-Nearest Neighbor (k-NN)

K-Nearest Neighbors (KNN) adalah metode pembelajaran mesin standar yang telah diperluas untuk usaha pertambangan data skala besar. Ide di baliknya adalah menggunakan sejumlah besar data pelatihan, di mana setiap titik data ditandai oleh seperangkat variabel. Secara konseptual, setiap titik dipetakan dalam ruang berdimensi tinggi, di mana setiap sumbu di ruang tersebut sesuai dengan variabel individual. Ketika kita memiliki titik data baru (pengujian), kita ingin mencari K tetangga terdekat yang paling dekat (yaitu, paling "mirip" dengan titik tersebut). Jumlah K biasanya dipilih sebagai akar kuadrat dari N, jumlah total titik dalam set data pelatihan. (Oleh karena itu, jika N adalah 400, $K = 20$).

KNN konseptualnya sederhana dan memiliki keuntungan sebagai nonparametrik. Metode ini dapat digunakan bahkan ketika variabel adalah kategorikal - meskipun jika Anda menggunakan variabel

numerik, sebaiknya memstandarisasi mereka (lihat Bagian 4.4 bab: Teknologi Inti: Pembelajaran Mesin dan Pemrosesan Bahasa Alami) untuk menghilangkan perbedaan skala. Tantangannya adalah ketika jumlah titik data sangat besar (misalnya, penjual buku online memiliki jutaan buku), metode khusus harus digunakan untuk dengan cepat mencari ruang dan menemukan item "paling mirip".

Biasanya, beberapa bentuk pra-pemrosesan digunakan, misalnya pengindeksan. Selain itu, daripada menggunakan semua titik data, titik data terpilih yang mewakili kluster individual ("prototipe") dapat digunakan untuk memfasilitasi pencarian terhadap item baru, dan kemudian tetangga yang sudah dihitung dari prototipe paling mirip juga ditampilkan. Demikian pula, upaya untuk mengurangi jumlah dimensi dengan metode seperti SVD/LSI dan kemudian memetakan titik data dalam ruang variabel yang direduksi dapat menghasilkan kenaikan kinerja yang signifikan.

Naïve Bayes (NB)

Naive Bayes adalah metode yang cocok untuk klasifikasi biner dan *multiclass*. Metode yang juga dikenal sebagai Naive Bayes Classifier ini menerapkan teknik *supervised* klasifikasi objek di masa depan dengan menetapkan label kelas ke *instance*/catatan menggunakan probabilitas bersyarat. Probabilitas bersyarat adalah ukuran peluang suatu peristiwa yang terjadi berdasarkan peristiwa lain yang telah (dengan asumsi, praduga, pernyataan, atau terbukti) terjadi.

Istilah *supervised* merujuk pada klasifikasi *training data* yang sudah diberi label dengan kelas. Misalnya, sebuah transaksi penipuan telah ditandai sebagai data transaksional. Kemudian, jika Anda ingin mengklasifikasikan transaksi di masa depan menjadi *fraudulent/non-fraudulent* (penipuan/non-penipuan), maka jenis klasifikasi itu akan disebut sebagai *supervised*.

Nah, model *machine learning* yang diterapkan pada program tersebut menggunakan teorema Bayes yang dirumuskan sebagai berikut:

$$P(A / B) = P(B | A)P(A)P(B)$$

Keterangan:

$P(A | B)$: Probabilitas A terjadi dengan bukti bahwa B telah terjadi
(probabilitas superior)

$P(B | A)$: Probabilitas B terjadi dengan bukti bahwa A telah terjadi

$P(A)$: Peluang terjadinya A

$P(B)$: Peluang terjadinya B

2. Review paper tersebut sesuai kaidah review paper

Jurnal pertama

JUDUL	Implementasi data mining menggunakan metode naive bayes dengan feature selection untuk prediksi kelulusan mahasiswa tepat waktu
JURNAL	Jurnal Ilmiah Sains dan Teknologi
VOLUME & HALAMAN	Volume 6 Nomor 1
TAHUN	Tahun 2022
PENULIS	Royan Habibie Sukarna¹, Yulian Ansori² Teknik Informatika, Institut Teknologi Tangerang Selatan Ilmu Komputer, Universitas Budi Luhur Email: royan@itts.ac.id , julianopec@gmail.com
REVIEWER	A120043 – Miftahul Jannah
TANGGAL	12 Mar. 2023
TUJUAN PENELITIAN	Maka penelitian ini bertujuan untuk memprediksi kelulusan dari mahasiswa yang tepat waktu sebagai penunjang pengambilan keputusan sebagai bagian dari upaya untuk meningkatkan AEE dengan implementasi dari data mining (Tekin, 2014). Sebagai upaya untuk meningkatkan persentase kelulusan mahasiswa tepat waktu adalah dengan

	menganalisis pola dalam database akademik, untuk memprediksi tingkat kelulusan tepat waktu yang sulit untuk dianalisis secara manual (Tahyudin et al., 2013).
SUBJEK PENELITIAN	Data Mining, Graduation Prediction, Naïve Bayes
METODE PENELITIAN	<p>METODE</p> <p>1. Data Mining</p> <p>Penggabungan bidang ilmu basis data, machine learning, statistika, information retrieval, dsb menghasilkan suatu ilmu baru yaitu Data Mining. Data Mining digunakan untuk menggali informasi tersembunyi dalam basis data yang merupakan bagian dari tahap proses Knowledge Discovery in Database (KDD). (Moertini, 2002).</p> <p>Data Mining dapat diartikan juga sebagai proses dalam eksplorasi dan analisis data yang dapat dilakukan dengan banyak metode yang memiliki kegunaan masing-masing. Atau dapat juga diartikan sebagai proses mengekstraksi informasi dari set data besar melalui penggunaan algoritma dan teknik yang diambil dari bidang statistik dan Manajemen Sistem Database (Ogunde & Ajibade, 2014).</p> <p>2. Naïve Bayes</p> <p>Algoritma Naive Bayes (NB) merupakan metode yang sederhana dalam klasifikasi berdasarkan teori probabilitas yang dikemukakan pertama kali oleh ilmuwan inggris bernama Thomas Bayes. Disebut naif karena menyederhanakan masalah yang bergantung pada dua asumsi penting (Osmanbegović & Suljić, 2012).</p> <p>Keuntungan dari klasifikasi Naïve Bayes adalah bahwa algoritma ini tidak membutuhkan data pelatihan dalam jumlah yang besar dalam proses</p>

	klasifikasi (Bhardwaj & Pal, 2011). Klasifikasi Naïve Bayes telah terbukti dapat diaplikasikan dalam situasi nyata dan kompleks.
LANGKAH-LANGKAH PERANCANGAN	Langkah-langkah Penelitian Langkah-langkah yang digunakan dalam penelitian ini penulis mengadopsi model CRISP-DM (Cross Standard Industries Process for Data Mining), dimana model ini terdapat 6 tahapan (Chapman et al., 2000),
HASIL PENELITIAN	Dalam paper ini, metode yang diusulkan yaitu menggunakan feature selection dengan algoritma Gain Ratio Attribute mendapatkan hasil yang baik dan mengalami peningkatan akurasi. Pada penelitian selanjutnya sangat dimungkinkan untuk mengimplementasikan feature selection kedalam algoritma klasifikasi yang lain, seperti KNN, C45, dsb.

Jurnal kedua

JUDUL	Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi menggunakan Algoritma K-Nearest Neighbor (K-NN)
JURNAL	Jurnal Sains dan Manajemen
VOLUME & HALAMAN	Vol 8 No. 1
TAHUN	Maret 2020
PENULIS	Retno Sari <i>Teknik Informatika, STMIK Nusa Mandiri Indonesia</i> <i>E-mail:retno.rnr@nusamandiri.ac.id</i>
REVIEWER	A120043 – Miftahul Jannah
TANGGAL	12 Mar. 2023

TUJUAN PENELITIAN	<p>Penelitian ini bertujuan untuk mengetahui akurasi analisis sentimen dengan menggunakan algoritma K-Nearest Neighbor yang mana review menggunakan teks berbahasa Indonesia. Kelebihan k-NN memiliki prinsip sederhana, bekerja berdasarkan jarak terpendek dari sampel uji dan sampel latih. Pada penelitian ini preprocessing yang digunakan yaitu tokenize, filter stopword dan bi-gram. Data yang digunakan terdiri dari 50 review positif dan 50 review negative. Eksperimen telah dilakukan dengan mengubah nilai k dan didapat akurasi yang terbesar dengan nilai k=7 dengan akurasi 77.01, precision 92.38, recall 61.56 dan nilai AUC 0.894.</p>
SUBJEK PENELITIAN	analisis sentiment; review objek wisata; k-NN
METODE PENELITIAN	<p>Metode</p> <p>c. Metode yang digunakan</p> <p>Pada penelitian ini menggunakan metode K-Nearest Neighbor (k-NN) untuk mengetahui akurasi yang dihasilkan.</p> <p>d. Eksperimen dan Pengujian Metode</p> <p>Eksperimen pada penelitian ini menggunakan Rapid Miner 5 dalam pengolahan data.</p> <p>e. Evaluasi dan Validasi Hasil</p> <p>Validasi dilakukan menggunakan fold cross validation, untuk melakukan pengukuran akurasi menggunakan confusion matrix dan kurva ROC untuk mengukur nilai AUC.</p>
LANGKAH-LANGKAH PERANCANGAN	<p>yang digunakan pada penelitian ini merupakan metode penelitian eksperimen, dengan tahapan sebagai berikut:</p> <p>a. Pengumpulan Data</p>

	<p>Penulis mengambil data dari situs review www.tripadvisor.com. Data yang diambil terdiri dari 50 review positif dan 50 review negative.</p> <p>b. Pengolahan Awal Data</p> <p>Preprocessing pada dataset yang dilakukan yaitu:</p> <p>1) Tokenize</p> <p>Tokenize merupakan tahap preprocessing dengan menghilangkan tanda baca maupun simbol apapun yang bukan huruf.</p> <p>2) Filter Stopword</p> <p>Pada tahap filter stopwords membuang kata-kata berupa kata sambung, imbuhan, akhiran. Seperti kata “dari”, “yang”, “di”, dan “ke”.</p> <p>3) N-gram</p> <p>Pada tahap ini dilakukan pemotongan karakter dengan bi-gram yaitu pemotongan string sebanyak 2 bagian.</p>
HASIL PENELITIAN	<p>Pada penelitian ini menggunakan data dari www.tripadvisor.com yang terdiri dari 50 data review positif dan 50 data review negative dengan teks berbahasa Indonesia. Data review tersebut dikelompokkan ke dalam folder positif untuk review positif dan dikelompokkan ke dalam folder negative untuk review negative.</p> <p>Proses pengolahan awal terdiri dari proses tokenize, filter stopwords dan n-gram. Pada proses tokenize semua review dihilangkan tanda baca maupun simbol apapun yang bukan huruf. Sehingga membentuk kumpulan kata yang memiliki arti. Pada tabel 1 dapat dilihat hasil dari proses tokenize.</p>

Jurnal ketiga

JUDUL	Integrated Library System untuk Meningkatkan Efektivitas Layanan Perpustakaan Dengan Menggunakan Metode Algoritma Apriori (Studi Kasus: Perpustakaan Kabupaten Cirebon)
JURNAL	JOINT (Journal of Information Technology)
VOLUME & HALAMAN	Vol. 03 No 01
TAHUN	Februari 2021
PENULIS	Alfi Nurfajiah ^{1*} , Tuti Hartati ² , Dita Rizki Amalia ³ 1,2,STMIK IKMI Cirebon 1alfinurfajiah06@gmail.com 2toohart2013@gmail.com 3ditarizkiamalia@gmail.com
REVIEWER	A120043 – Miftahul Jannah
TANGGAL	12 Mar. 2023
TUJUAN PENELITIAN	Tujuan dari penelitian ini untuk menentukan metode peminjaman buku yang direkomendasikan oleh Dinas Perpustakaan dan Kearsipan Daerah (DPAD), sehingga pustakawan dapat lebih mudah menentukan buku yang ingin mereka rekomendasikan kepada pembaca. Dalam penelitian ini menggunakan metode algoritma apriori, dimana metode tersebut menentukan itemset dari transaksi peminjaman buku, kemudian mengelompokkan menjadi tabel itemset1, tabel itemset2, tabel nilai dukungan (nilai confidence), tabel nilai kepercayaan dan memperoleh mode peminjaman buku
SUBJEK PENELITIAN	Perpustakaan Kabupaten Cirebon, Algoritma Apriori
METODE PENELITIAN	Analisis Kualitas Layanan Sistem Informasi Inlis lite Menggunakan Metode LibQual, Jurnal Sains, Teknologi dan Industri, Vol. 13, No. 2, Fakultas Sains dan Teknologi, UIN Suska Riau, Juni 2016, Nurmaini Dalimunthe, Dewi Sartika, menjelaskan

	<p>bahwa INLIS (Integrated Library System) LITE merupakan sebuah sistem informasi Perpustakaan yang terintegrasi berbasis localhost yang diterapkan pada Badan Perpustakaan dan Arsip Kabupaten Cirebon pada tahun 2020. Sistem ini untuk menyediakan informasi guna mendukung operasional, manajemen, dan fungsi pengambilan keputusan dalam Perpustakaan. Fitur layanan yang ada didalamnya yaitu Backoffice, keanggotaan, pendaftaran anggota dan checkpoint. Metode libqual adalah pengembangan dari servqual (service quality) yang digunakan untuk mengukur tingkat kepuasan pemustaka terhadap kualitas pelayanan perpustakaan. Libqual (library quality) merupakan salah satu panduan layanan yang digunakan Perpustakaan untuk mengumpulkan, memetakan, memahami dan bertindak atas pendapat pengguna Perpustakaan terhadap kualitas layanan Perpustakaan[1].</p>
LANGKAH-LANGKAH PERANCANGAN	<p>Teknik Pengambilan Data</p> <ol style="list-style-type: none"> 1. Observasi 2. Kuesioner 3. Wawancara 4. Studi Pustaka 5. Survey
HASIL PENELITIAN	<p>Hasil dan pembahasan, pembahasan pada tahap ini menjelaskan hasil dari proses data mining yang dilakukan dengan menggunakan algoritma apriori. Kesimpulan dan saran, membuat kesimpulan dari hasil penelitian dan memberikan saran untuk pihak Perpustakaan agar dapat menjadi lebih baik.</p>

Jurnal keempat

JUDUL	Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree
JURNAL	Scientific Journal of Informatics
VOLUME & HALAMAN	Vol. 3, No. 1,
TAHUN	Mei 2016
PENULIS	Ari Muzakir¹, Rika Anisa Wulandari² ^{1,2} Jurusan Informatika, F. ILKOM, Universitas Bina Darma Palembang Email: 1arimuzakir@binadarma.ac.id, 2ranisa.wulandari@yahoo.com
REVIEWER	A120043 – Miftahul Jannah
TANGGAL	12 Mar. 2023
TUJUAN PENELITIAN	Perkembangan teknologi informasi saat ini sangat berguna bagi semua kalangan masyarakat. Sekarang ini teknologi informasi telah menjadi salah satu kebutuhan yang sangat penting dalam kehidupan sehari-hari. Teknologi informasi dapat mempermudah manusia dalam mengerjakan pekerjaannya. Sehingga, teknologi informasi dapat diimplementasikan dalam berbagai bidang, termasuk dunia kedokteran dan kesehatan. Kesehatan merupakan aspek terpenting didalam setiap kehidupan, khususnya bagi ibu hamil. Hipertensi merupakan salah satu masalah kesehatan yang sering muncul selama kehamilan dan dapat menimbulkan komplikasi pada 2–3% kehamilan. Hipertensi pada kehamilan juga dapat berlanjut menjadi preeklamsia dan eklamsia yang dapat menyebabkan kematian pada ibu maupun janin
SUBJEK PENELITIAN	Data mining, Decision tree, C4.5 algorithms, Prediction, Pregnancy
METODE PENELITIAN	metode tertentu. Teknik, metode atau algoritma data mining sangat bervariasi. Pemilihan metode atau

	<p>algoritma yang tepat sangat bergantung pada tujuan dan proses Knowledge Discovery in Database (KDD). Adapun teknik yang digunakan pada proses data mining adalah dengan teknik decision tree dengan menggunakan algoritma C4.5.</p> <p>2.3.1 Decision Tree</p> <p>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. Berikut adalah struktur decision tree seperti yang ditunjukkan pada Gambar 4.</p>
LANGKAH-LANGKAH PERANCANGAN	<p>Salah satu metode yang dapat digunakan untuk memprediksi penyakit hipertensi adalah dengan menggunakan data mining. Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang</p>

	<p>selama ini tidak diketahui secara manual [2]. Metode untuk menganalisis data dalam penerapan data mining ini adalah Knowledge Discovery in Database (KDD) yang terdiri dari beberapa tahapan, yaitu Cleaning and Integration, Selection and Transformation, Data mining, dan Evaluation and Interpretation [1].</p> <p>Adapun teknik yang digunakan dalam data mining adalah decision tree (pohon keputusan) dengan menggunakan algoritma C4.5. Dengan menggunakan teknik decision tree menggunakan algoritma C4.5, akan menghasilkan pola atau pattern, yaitu berupa rules yang dapat memprediksi penyakit hipertensi dalam kehamilan.</p>
HASIL PENELITIAN	<p>Implementasi data mining dengan teknik decision tree menggunakan algoritma C4.5 dapat menghasilkan informasi berupa prediksi penyakit hipertensi dalam kehamilan, dimana dari data training yang digunakan dengan jumlah 286 instance dapat dibangun sebuah decision tree yang menghasilkan rules yang bisa digunakan untuk memprediksi penyakit hipertensi dalam kehamilan.</p> <p>Dari decision tree yang dibangun, menunjukkan bahwa atribut yang menjadi faktor pendukung seorang ibu hamil bisa menderita penyakit hipertensi dalam kehamilannya, yaitu berdasarkan usia, berat badan, riwayat hipertensi, dan paritas.</p> <p>Setelah mendapatkan decision tree dan rules yang dapat memprediksi penyakit hipertensi dalam kehamilan, dilakukan evaluasi dengan supplied test set menggunakan WEKA dihasilkan kesalahan (error) 7.3427% dan tingkat akurasi 92.6573%. Data training yang berjumlah 286 instances, hal ini</p>

	menunjukkan bahwa terdapat 265 instances yang akurat dan 21 instances yang error atau prediksinya salah.
--	--

Jurnal kelima

JUDUL	KLASIFIKASI PEMINJAMAN NASABAH BANK MENGGUNAKAN METODE NEURAL NETWORK
JURNAL	Jurnal PILAR Nusa Mandiri
VOLUME & HALAMAN	Vol. 15, No. 2
TAHUN	September 2019
PENULIS	Nur Hadiano ¹ ; Hafifah Bella Novitasari ² ; Ami Rahmawati ³ Ilmu Komputer STMIK Nusa Mandiri Jakarta www.nusamandiri.ac.id 1dion.adiyanto@gmail.com; 2hafifah.bella@gmail.com; 3amirahmawati095@gmail.com
REVIEWER	A120043 – Miftahul Jannah
TANGGAL	12 Mar. 2023
TUJUAN PENELITIAN	Penelitian ini bertujuan untuk mengklasifikasi nasabah yang layak untuk mendapatkan pinjaman dengan memperhitungkan parameter yang ada seperti usia, jumlah pendapatan, jumlah keluarga, rata-rata pengeluaran perbulan tingkat pendidikan dan lainnya. Penelitian ini menggunakan metode klasifikasi data mining dengan model neural network, untuk menilai akurasi pengolahan data
SUBJEK PENELITIAN	Pinjaman, Klasifikasi, Neural Network, Data Mining, Rapid Miner, Backpropagation.

METODE PENELITIAN	<p>Penelitian ini menggunakan metode klasifikasi data mining dengan model neural network, untuk menilai akurasi pengolahan data menggunakan rapid miner kemudian dilanjutkan dengan pengukuran menggunakan confusion matrix, kurva ROC. Hasil algoritma neural network setelah melalui pengujian confusion matrix, kurva ROC menunjukkan nilai akurasi yang sangat tinggi, dan nilai dominan AUC dan algoritma. Nilai akurasi adalah 98,24% dengan AUC sebesar 0,979</p>
LANGKAH-LANGKAH PERANCANGAN	<p>Setelah pada tahap proses pengumpulan data dan pemilihan atribut yang akan digunakan di dalam dataset maka langkah selanjutnya adalah kita harus memasukkan datanya terlebih dahulu dengan fungsi read excel kedalam tools rapid miner lalu pembuatan permodelan neural network menggunakan data sebanyak 5000, tahapan proses dimulai dengan tahap permodelan kemudian memasuki tahap learning guna memproses datatraining dengan ujicoba struktur, training cycles, learning rate, momentum kemudian diakhiri dengan testing untuk melihat hasil.</p> <p>1. Proses Permodelan</p> <p>Proses ini menggunakan cross validation untuk melakukan pengujian model setelah sebelumnya Pada gambar diatas dilakukan proses pengujian model yang telah dibaca menggunakan operator read excel, data yang telah dibaca tersebut kemudian dimasukkan kedalam operator cross validation, dalam penelitian ini cross validation yang digunakan sebesar 10 fold validation.</p> <p>2. Proses Training</p>

	<p>Proses selanjutnya yang dilakukan adalah proses training, proses ini dilakukan di dalam operator cross validation dengan menggunakan algoritma neural network, dengan algoritma ini maka data akan dibagi menjadi 2, bagian pertama digunakan sebagai data training dan bagian kedua digunakan sebagai data testing, seperti pada gambar 7 dibawah ini. Hasil Testing</p> <p>Hasil dari training dan juga testing data terhadap 5000 data nasabah Bank yang mengajukan pinjaman dapat kita lihat pada Performance Vector seperti pada gambar 9 dibawah ini.</p> <p>Hasil testing menggunakan algoritma neural network menunjukkan performance accuracy yang didapat sebesar 98.24%.</p> <p>5. Hasil Uji ROC Curve</p> <p>Hasil dari training dan juga testing data menggunakan neural network menampilkan grafik ROC dengan nilai AUC (Area Under Curve) sebesar , hasil tersebut dapat dilihat seperti pada gambar 10 dibawah ini.</p>
HASIL PENELITIAN	<p>Pada bagian ini akan membahas hasil implementasi penggunaan Neural Network pada penentuan klasifikasi peminjaman nasabah pada sebuah bank dengan menggunakan metode backpropagation, Analisa difokuskan pada pengaruh struktur neural network, tingkat pembelajaran (learning rate) neural network, dan momentum neural network, dalam penelitian ini menggunakan dataset sebanyak 5000 data.</p> <p>Dalam penelitian untuk mendapatkan struktur neural network yang optimal kami melakukan beberapa ujicoba perubahan struktur dengan tingkat</p>

	<p>pembelajaran (learning rate), training cycles dan momentum yang sama yakni learning rate = 0.0.1, training cycles = 500 dan momentum = 0.1, pertama kali kami mencoba struktur 12 – 7 – 3 – 1 artinya struktur tersebut memiliki 12 input, 7 neuron pertama yang tersembunyi, 3 neuron kedua yang tersembunyi dan 1 output, kemudian kami mencoba beberapa struktur yang lain seperti 12 – 20 – 8 – 1 , 12 – 15 – 8 – 1 , 12 – 10 – 5 – 1 , 12 – 7 – 3 – 1, dari semua struktur yang dicoba tidak terdapat perbedaan yang signifikan hanya selisih 0 (nol) koma sekian, akan tetapi dari seluruh struktur yang sudah kami coba kami menemukan bahwa struktur 12 – 15 – 8 – 1 merupakan struktur yang paling adaptif dengan akurasi 98.24 %, struktur 12 – 15 – 8 – 1 adalah seperti pada gambar 4. dibawah ini.</p>
--	--