

BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Tabel 2.1 Penelitian Terkait

No	Judul	Masalah	Metode Yang Digunakan	Hasil	Tahun
1	PENGALAMAN PSIKOSOSIAL FIRST AID (DEPRESI POSTPARTUM) PADA IBU PRIMIPARA DENGAN RIWAYAT SECTIO CAESAREA	Gangguan ini merefleksikan disregulasi psikologikal yang merupakan tanda dari gejala-gejala depresi mayor. Depresi postpartum juga dapat disebabkan oleh pengaruh dari jenis persalinan.	instrumen Beck Depression Inventory (BDI-II). purposive sampling	Hasil penelitian terdapat trauma pada P1 dan P2 karena proses persalinan, terkait penanganan P1 dengan menonton terutama yang memancing gelak tawa supaya mood kembali normal, selain itu juga dengan mengonsumsi makanan yang disukai dan juga menyuruh teman atau saudara terdekat menemani untuk sekedar bercerita sedangkan P2 dengan melihat anaknya. Depresi postpartum dapat terjadi karena proses persalinan, dukungan keluarga dan lingkungan sekitar.[4]	2022
2	Implementasi Teknik Data Mining untuk mendeteksi Gangguan Psikologis Pasca Melahirkan	Salah satu gangguan psikologis terjadi pada ibu pasca melahirkan, sekitar 10-15% ibu pasca melahirkan mengalami gangguan psikologis seperti kecemasan dan depresi, tingginya angka tersebut menjadi kekhawatiran di lingkungan masyarakat, terutama keluarga.	data mining, naïve Bayes,	Dari hasil evaluasi model, dapat disimpulkan bahwa model Naive Bayes menunjukkan performa yang baik dalam mendeteksi gangguan psikologis pasca melahirkan. Nilai-nilai evaluasi menggambarkan kemampuan model dalam membedakan antara kelas positif dan negatif, dengan nilai AUC sekitar 0.878. Akurasi model sekitar	2023

				0.819 mengindikasikan kemampuannya dalam memprediksi dengan benar sekitar 81.9% dari kasus yang diuji.[5]	
3	Penerapan Sistem Pakar Berbasis Android Dengan Metode Decision Tree Untuk Memprediksi Postpartum Haemorrhage Pada Wanita Hamil	Postpartum Haemorrhage (perdarahan pasca melahirkan) menjadi faktor kematian ibu sebesar 25% di dunia . Di negaranegara berkembang, faktor Haemorrhage atau perdarahan mewakili kematian ibu 661.000 jiwa di seluruh dunia dengan 27,1% pada range 19,9 – 36,2 kematian ibu, lalu lebih dari dua pertiga kematian haemorrhage (perdarahan) diklasifikasikan sebagai postpartum (pasca melahirkan) 19,7% pada range 12,9 – 28,9 dengan kematian ibu 480.000 jiwa di seluruh dunia dan 479.000 jiwa berasal dari Negara berkembang .	Decision Tree	keterangan yang dapat dilihat pada hasil di pohon keputusan, dari keterangan PPH dan NO PPH menjadi keterangan PPH, NO PPH dan Potential PPH. Dari Pengujian menggunakan RapidMiner pada data Training mendapat Accuracy 93,91%, dan pada data Testing mendapat Accuracy 94,86%. Kemudian telah dibuatnya sebuah sistem pakar berbasis Android dalam membantu meringankan kinerja bidan dalam mengarahkan sejak dini untuk konsultasi ke dokter setelah melihat hasil dari kemungkinan terjadinya postpartum haemorrhage yang ditampilkan oleh sistem pakar, hingga dokter beserta tim tenaga medis dapat menyediakan tindakan lanjut untuk para wanita hamil yang memiliki kemungkinan mengalami postpartum haemorrhage.	2021
4	Implementation Data Mining using Decision Tree Method- Algorithm C4.5 for Postpartum Depression Diagnosi	Postpartum depression is a serious problem that needs to be addressed because it has negative effects on family, child welfare, cognitive, and mother child interactions. Diagnosis is done based on psychological condition, blood pressure, respiration, body temperature, and classification data extract by	Decision Tree Method- Algorithm C4.5	Results of this study in the form of an online information system that can identify the level of depression more quickly and precisely. The results showed the greatest gain on the psychological variables of 0.57 node 1, blood pressure 0.54 node 2, body temperature 0,54	2022

		decision tree C4.5 algorithm method.		node 3, means that the three variables are more influential on the condition of depressed patients, and should be given priority treatment. Test results from 50 patients with 50 examinations showed 62% prevalence, 65.62% sensitivity, specificity 77.77%, negative predictive value of 56%, and positive predictive value 84%	
5	Seleksi Atribut Pada Diagnosis Penyakit Liver Menggunakan Decision Tree Dengan Algoritma Genetika	Penyakit Liver merupakan penyakit krusial pada organ hati di mana penyakit tersebut diakibatkan oleh inflamasi pada hati yang disebabkan infeksi virus, bahan beracun, atau bakteri sehingga hati kesulitan untuk berfungsi sebagaimana mestinya [1]. Berdasarkan data yang dikeluarkan oleh WHO (World Health Organization) bahwa setidaknya hampir 1.2 juta orang per tahun di kawasan Asia Tenggara dan kawasan Afrika mengalami kematian akibat terserang penyakit ini [2]. Agar penyakit tersebut tidak semakin parah, maka dari itu diperlukan diagnosis sejak dini agar penderita mendapatkan penanganan sesegera mungkin.	Decision Tree Dengan Algoritma	Dalam hasil evaluasi, aturan yang dihasilkan oleh optimasi adalah lebih sederhana yaitu sebanyak 32 aturan jika dibandingkan tanpa optimasi yang lebih kompleks, yaitu 145 aturan. Kemudian pada evaluasi akurasi, aturan dengan optimasi menghasilkan hasil yang lebih baik akurasi sebesar 70,7% jika dibandingkan akurasi aturan tanpa optimalisasi sebesar 68,9%. [6]	2018

2.2 Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan. Atau dapat diartikan Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.[7]

2.3 Tahapan Data Mining

Istilah *Knowledge Discovery in Database* (KDD) dan data mining seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain, dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining

Sesuai yang tercantum dalam buku “*Advances in Knowledge Discovery dan Data mining*” terdapat definisi sebagai berikut: Knowledge discovery (data mining) in databases (KDD) adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah (valid), baru (novel), dapat bermanfaat (*potentially usefull*), dapat dimengerti (*ultimately understandable*). (Vapnik, 1998). Istilah data mining dan *Knowledge Discovery In Databases* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis besar dapat dijelaskan sebagai berikut :[8]

A. Data Selection

Pemilihan (seleksi) data dan sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

B. Pre-processing/Cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

C. Transformation Coding

Proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

D. Data mining

Proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi.

E. Interpretation / evaluation

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.



Gambar 2.1 Tahapan Data Mining

2.4 Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari ID3 ini dapat mengklasifikasikan data dengan metode pohon keputusan yang memiliki kelebihan dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan, dan tercepat diantara algoritma-algoritma yang menggunakan memori utama di komputer. Algoritma C4.5 menggunakan kriteria gain dalam menentukan fitur yang menjadi pemecah node pada pohon yang diinduksi. Berikut ini formulasi tahapan pada Algoritma pada persamaan berikut[9] :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{S_i}{S} * Entropy(S_i)$$

Keterangan :

S : Himpunan Kasus

A : Atribut

N : Jumlah partisi atribut A

|S_i| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Sedangkan perhitungan nilai Entropy dapat dilihat pada persamaan berikut :

$$Entropy(A) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Keterangan :

S : Himpunan Kasus
 A : Fitur
 n : Jumlah partisi S
 pi : Proporsi dari Si terhadap S

Kriteria yang paling banyak digunakan sebagai pemecah dalam Algoritma C4.5 adalah Gain Ratio (Rasio Gain), dengan persamaan berikut :

$$RasioGain (s, j) = \frac{Gain (s, j)}{SplitInfo(s, j)}$$

Keterangan :

Gain(s,j) : Nilai Information

Gain SplitInfo(s,j) : Nilai Splitinfo

Dari persamaan menyatakan nilai rasio gain pada fitur ke-j. SplitInfo(s,j) didapat dari persamaan 4 berikut :

$$SplitInfo(s, j) = \sum_{i=1}^k p(v \log_2 p(vi|s))$$

Keterangan :

s : Data yang digunakan untuk training / testing

j : Atribut

vi : jumlah data untuk atribut ke-i

2.5 *K-Nearest Neighbors* (KNN)

K-Nearest Neighbor (KNN) termasuk algoritma supervised learning dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Nanti kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi. KNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing. Algoritma *K-Nearest Neighbor* adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. KNearest Neighbor adalah pendekatan untuk mencari kasus dengan

menghitung kedekatan antara kasus baru dan kasus lama yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada.[10]

Algoritma KNN adalah algoritma klasifikasi atau regresi nonparametrik dalam bidang pengenalan pola. Pada subbagian ini, kami secara singkat memperkenalkan algoritma klasifikasi KNN. Asumsikan ada beberapa data pelatihan yang memiliki beberapa atribut dan label. Selanjutnya ada kelompok data testing yang hanya memiliki beberapa atribut saja tanpa label. Tujuan dari algoritma klasifikasi KNN adalah untuk mendapatkan label dari data pengujian. Proses spesifiknya adalah sebagai berikut:

- $(x_i, y_i), i = 1, 2, \dots, N$
- x_i adalah data pelatihan dalam R^n
- y_i adalah kelas yang sesuai dari data x_i , dan $y_i \in \{c_j, j = 1, 2, \dots, M\}$
- $\text{dist}(x - x_i) = \|x - x_i\|$

Algoritma KNN klasik dapat digunakan untuk prediksi numerik, yaitu mengembalikan nilai prediksi sebenarnya menurut tuple yang tidak diketahui. Strategi pemasaran harga harus memahami dasar objektif penetapan harga; melakukan riset pasar; menganalisis secara objektif kebijakan nasional, lingkungan pasar, nilai proyek pembangunan itu sendiri, hubungan penawaran dan permintaan pasar real estat, situasi persaingan pasar, psikologi pembeli rumah, dan faktor lainnya; pilih metode penetapan harga ilmiah; dan mengadopsi strategi yang sesuai.[11]

2.6 *Postpartum Depression*

Postpartum depression (PPD) adalah salah satu jenis depresi yang dialami setelah proses persalinan dengan onset pada satu bulan pertama setelah melahirkan, yaitu pada minggu ke-2 hingga ke-6 dan terjadi pada 13% ibu-ibu setelah melahirkan. Penegakan diagnosis gangguan depresi ini memerlukan waktu sekurang-sekurangnya dua minggu. PPD biasanya diawali dengan sindroma gangguan afek ringan yaitu *postpartum blues* (*baby blues syndrome*). Apabila PPD tidak tertangani dengan baik, maka gejalanya dapat berkembang menjadi gejala gangguan yang lebih serius yaitu *postpartum psychosis*. [12]

Depresi postpartum (Postpartum Depression) adalah salah satu bentuk depresi yang dialami ibu setelah melahirkan bayi pertama dan berlangsung pada tahun pertama

setelah kelahiran bayi. Hal ini disebabkan karena periode tersebut merupakan periode transisi kehidupan baru yang cukup membuat stres, dimana ibu harus beradaptasi dengan perubahan fisik, psikologis dan sosial yang dialaminya karena melahirkan dan mulai merawat bayi. Namun tidak semua ibu mampu melakukan adaptasi dan mengatasi stressor tersebut sehingga timbul keluhan-keluhan antara lain berupa stres, cemas dan depresi.[13]

Menurut Diagnostic and Statistical Manual of Mental Disorder-IV, ada 3 jenis bentuk depresi pasca melahirkan, yaitu :

1. Postpartum Blues yang merupakan gangguan mood yang bersifat sementara.
2. Depresi Postpartum tanpa gambaran psychosis yang lebih berat dari Postpartum Blues.
3. Postpartum Psychosis, yaitu depresi berat berupa gangguan proses pikir yang dapat mengancam keselamatan jiwa ibu dan bayinya sehingga memerlukan bantuan psikiater.

2.7 Rapid Miner

Rapid Miner adalah platform perangkat lunak yang kuat untuk ilmu data dan pembelajaran mesin. Ini menyediakan beragam alat untuk persiapan data, pemodelan, evaluasi, dan implementasi. Rapid Miner dirancang untuk mudah digunakan dan memungkinkan pengguna untuk dengan mudah membangun dan menguji berbagai model, bahkan tanpa pengalaman pemrograman.

Rapid Miner menawarkan antarmuka drag-and-drop yang memungkinkan pengguna untuk membangun alur kerja untuk memproses dan menganalisis data. Ini mendukung beragam sumber data, termasuk file datar, basis data, dan platform big data seperti hadoop dan spark. Perangkat lunak ini juga mencakup beragam operator yang sudah dibangun yang merupakan blok bangunan dari alur kerja, yang mencakup semua tahap proses data mining, seperti pembersihan data, pemilihan fitur, dan pemodelan.[14]

2.8 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) merupakan teknik optimasi yang berbasis populasi yang di usulkan oleh Eberhart dan Kennedy pada tahun 1995. Metode ini terinspirasi pada perilaku sosial sekawan burung dan ikan. Dalam mencari solusi yang

optimal, partikel tersebut bergerak pada arah yang terbaik sebelumnya, posisi terbaik secara global.[9]

Algoritma PSO dapat digunakan pada berbagai masalah optimasi baik kontinyu maupun diskrit, linier maupun nonlinier. PSO memodelkan aktivitas pencarian solusi terbaik dalam suatu ruang solusi sebagai aktivitas terbangnya kelompok partikel dalam suatu ruang solusi tersebut. Posisi partikel dalam ruang solusi tersebut merupakan kandidat solusi yang berisi variabel-variabel optimasi. Setiap posisi tersebut akan dikaitkan dengan sebuah nilai yang disebut nilai objektif atau nilai fitness yang dihitung berdasarkan fungsi objektif dari masalah optimasi yang akan diselesaikan. [15]

2.9 Confusion Matrix

Perhitungan Akurasi untuk big data pada konsep data mining digunakan untuk mendapatkan hasil klasifikasi prediksi dan hasil klasifikasi aktual yang telah didapat oleh sistem klasifikasi yang disebut Confusion Matrix . Sebuah data matrix digunakan untuk evaluasi performa dari sebuah sistem.. untuk menghitung akurasi dapat melakukan sesuai persamaan 7, untuk precision dapat dilihat pada persamaan 8, dan untuk recall dapat dilihat pada persamaan 7.[16]

1. Akurasi merupakan Merupakan rasio prediksi Benar (positif dan negatif) dengan keseluruhan data. Akurasi = $(TP+TN)/(TN+FP+FN+TP)$ (8)
2. Precision Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Precision = $TP/(FP+TP)$ (9)
3. Recall Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif $Recall = \frac{TP}{FN+TP}$ (10)

Keterangan :

- TN (True Negative) :Jumlah data yang teridentifikasi denan benar sebagai kelas negative
- FP (False Negative) : Jumlah data negatif yang teridentifikasi sebagai kelas positif
- FN (True Positive) : Jumah data posiiif yang teridentifikasi sebagai kelas negatif

- TP (True Positive) : Jumlah data yang teridentifikasi dengan benar sebagai kelas positif.

2.10 Klasifikasi

Klasifikasi merupakan tatanan yang sangat penting dalam menambang data komunitas. Klasifikasi adalah teknik penambangan data prediktif yang menggunakan hasil yang diketahui dari kumpulan data yang berbeda untuk membuat prediksi tentang data nilai. Masalah dengan akurasi banyak algoritma klasifikasi adalah bahwa informasi diketahui hilang saat memproses data yang tidak seimbang, misalnya ketika distribusi sampel antar kelas sangat miring.[17]