

BAB II STUDI PUSTAKA

2.1 Penelitian Terkait

Penelitian ini tidak lepas dari hasil penelitian – penelitian yang sudah dilakukan oleh penelitian sebelumnya sebagai referensi penunjang dalam penelitian ini. Adapun penelitian yang dijadikan acuan tidak lepas dari tema di atas yaitu :

Penelitian pertama dilakukan oleh (Sihotang, 2017) dengan judul” **PERANCANGAN APLIKASI SISTEM PAKAR DIAGNOSA DIABETES DENGAN METODE BAYES**” Deteksi secara dini dapat meminimalisasi resiko kematian penderita Diabetes Mellitus serta buat penanganan awal penyakit ini. Proses deteksi yang ada selama ini masih manual, hal ini bergantung di ahli penyakit organ pada tubuh manusia yg jumlahnya tergolong terbatas serta biayadiagnosa yang relative mahal. Akan tetapi, banyak warga yang belum memiliki pencerahan yang tinggi akan bahaya penyakit ini dan juga tidak memiliki pengetahuan dasar tentang penyakit ini dan mengalami keterbatasan waktu sehingga sporadis melakukan konsultasi kepada dokter. Oleh karena itu, dibutuhkan sebuah sistem ahli yang dapat menggantikan peran seorang dokter serta menyampaikan edukasi pengetahuan-pengetahuan awam tentang diabetes mellitus pada rakyat. Tujuan asal penelitian ini ialah merancang serta menciptakan sebuah sistem ahli yang bisa mendeteksi kemungkinan menderita diabetes mellitus. Berdasarkan gejala-gejala yg telah dimasukkan dan menyampaikan rekomendasi berupa info serta solusi terhadap penyakit tersebut. Salah satu metode yg bisa dipergunakan pada perhitungan probabilitasnya artinya metode Bayes. Metode ini artinya sebuah pendekatan buat sebuah ketidak-tentuan yang diukur menggunakan probabilitas. Sesuai hasil analisa, sistem ini bisa menangani proses konsultasi pengguna menggunakan tingkat akurasi yang relatif tinggi.

Penelitian kedua dilakukan oleh (Riadi, 2017) dengan judul **“PENERAPAN METODE CERTAINTY FACTOR UNTUK SISTEM PAKAR DIAGNOSA PENYAKIT DIABETES MELITUS PADA RSUD BUMI PANUA KABUPATEN POHUWATO”**. Diabetes Melitus (DM) atau biasa dianggap diabetes adalah penyakit gangguan metabolik menahun akibat pankreas tidak menghasilkan cukup insulin atau tubuh tidak bisa memakai insulin yang diproduksi secara efektif. Penderita Diabetes Melitus pada Kabupaten Pohuwato mengalami peningkatan sebanyak 8,5% setiap Tahun. Sistem pakar adalah acara personal komputer yang menirukan penalaran seorang ahli dengan keahlian pada suatu daerah pengetahuan eksklusif. Sistem pakar mencoba mencari solusi, memberikan saran atau konklusi yang konsisten terhadap perseteruan yang ditemukannya. Penelitian ini akan dibuat memakai aplikasi Dreamweaver dan bahasa pemrograman PHP, dan database MySQL. harapan penulis, sistem ini bisa membantu warga pada mendiagnosa penyakit Diabetes Melitus. Melalui software ini, pengguna bisa melakukan konsultasi menggunakan sistem layaknya berkonsultasi dengan seorang ahli untuk mendiagnosa gejala yang terjadi pada pengguna dan menemukan solusi atas pertarungan yang dihadapi. yang akan terjadi pengujian sistem diperoleh nilai Cyclomatic complexity = lima menggunakan jumlah Region (R)= lima, Node (N)= 10, Edge (E)=13 Predicate Node (P) = 4.

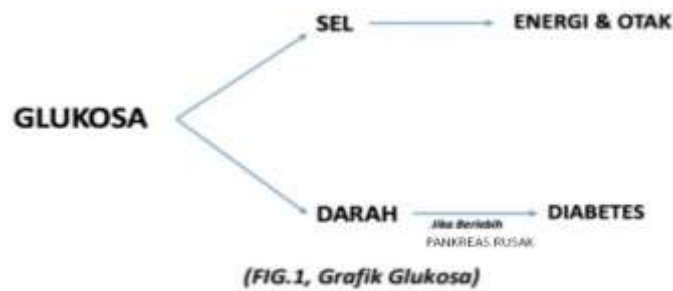
Peneliti ketiga yaitu di lakukan oleh (Jasri 1), 2018) dengan judul **“PENERAPAN METODE MAMDANI UNTUK SISTEM PENDUKUNG KEPUTUSAN PENENTUAN GOLONGAN OBAT SESUAI DENGAN PENYAKIT DIABETE”** Diabetes Meletus artinya yang tidak sesuai dengan dosis penyakit mengakibatkan tekanan darah glukosa menjadi tidak normal sebagai akibatnya sulitnya penderita Diabetes Meletus untuk menerima obat yang sesuai. untuk itu pertarungan akan diselesaikan dengan metode mamdani menggunakan beberapa variable yaitu variable input Kadar Glukosa Darah variabel hasil Golongan Obat. menggunakan penerapan variable akan mampu memilih dosis yang

sinkron dengan jenis penyakit sampai kadar glukosa darah supaya permanen normal. Penggunaan software ini diperuntukan buat apoteker hingga mampu dimanfaatkan Bila pasien benar benar membutuhkan takaran yang sinkron Jika dokter tidak terdapat di daerah. sesuai berasal ketiga penelitian di atas dapat pada simpulkan bahwa perbandingan yang pada jadikan alasan peneliti terletak pada metode penelitian guna memperoleh hasil yang sesuai dengan yang di harapkan. pada peneliti ketika ini mengajukan topik yg tidak sinkron menggunakan metode yang berbeda pula. Peneliti ini akan menggunakan metode decision tree C4.5 dalam mengklasifikasi penyakit diabetes. Metode decision tree c4.5 ialah metode yang di pakai buat melakukan klasifikasi terhadap suatu objek berdasarkan data training yang mempunyai jeda paling dekat. Tujuan asal algoritma decision tree c4.5 artinya mengklasifikasikan objek baru berdasarkan dari atribut dan sampel asal data pembinaan. menggunakan menerapkan metode decision tree c4.5 pada harapkan dapat di temukan aturan klasifikasi yang bisa di gunakan untuk membangun suatu sistem pendukung keputusan untuk mengidentifikasi seseorang terkena penyakit diabetes atau tidak.

2.2 Landasan Teori

2.2.1 Diabetes

Diabetes adalah meningkatnya gula darah akibat terlalu banyak mengonsumsi makanan karbohidrat seperti nasi putih dan gula, karbohidrat akan menjadi glukosa lalu glukosa akan masuk ke dalam 2 tempat tubuh manusia yaitu SEL dan DARAH. Apabila glukosa masuk ke SEL, glukosa tersebut akan menjadi Energi. Akan tetapi jika glukosa masuk terlalu banyak ke Darah maka akan menyebabkan gula darah tinggi. (Suswanto)



Gambar 2. 1 Grafik Glukosa

2.2.2 Prediksi

Prediksi merupakan proses dimana memperkirakan sesuatu yang paling mungkin akan terjadi pada masa yang akan tiba sesuai isu masa lalu atau kini yang di miliki, agar kesalahan antara sesuatu yang terjadi menggunakan yang akan terjadi dapat di perkecil. Prediksi adalah suatu proses memperkirakan hal yang paling mungkin terjadi di masa depan berdasarkan informasi masalalu dan kini, yang di miliki prediksi tidak harus memberikan jawaban secara tepat akan tetapi mencari jawaban sedekat mungkin apa yg akan terjadi (Ginting, Kusriani, & Emha Taufiq, 2020) Prediksi sama halnya dengan ramalah atau perkiraan. Menurut kamus besar bahasa Indonesia prediksi merupakan ramalan atau memperkirakan nilai pada masa yang akan tiba dengan menggunakan data masa lalu. prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu artinya input bagi proses perencanaan serta pengambilan keputusan (Ginting, Kusriani, & Emha Taufiq, 2020).

Peramal (*forecasting*) adalah prosedur untuk membuat informasi factual tentang situasi social masa depan atas dasar informasi yang telah ada tentang masalah kebijakan. Ramalan mempunyai tiga bentuk utama: proyeksi, prediksi, dan perkiraan.

1. Suatu proyeksi artinya ramalan yang di dasarkan pada ekstrapolasi atas kesamaan masa lalu juga masa kini ke masa depan. Proyeksi membentuk pernyataan yang tegas sesuai argument yang dapat di peroleh asal metode eksklusif dan kasus yang paralel.
2. Sebuah prediksi merupakan ramalan yang di dasarkan pada perkiraan teoritik yang tegas. asumsi ini dapat berbentuk aturan teoretis (contohnya aturan berkurangnya nilai uang) proposisi teoritis (misalkan proposisi bahwa pecahan masyarakat sipil di akibatkan sang kesenjangan antara harapan serta kemampuan) atau analogi antara pertumbuhan organisasi pemerintah dengan pertumbuhan organisme menggunakan biologis).

Suatu perkiraan (conjecture) adalah ramalan yang di dasarkan pada penilaian yang informative atau evaluasi pakar wacana situasi masyarakat masa depan.

Tujuan di adakannya ramalan adalah untuk memperoleh informasi tentang perubahan dimasa akan datang yang mempengaruhi terhadap implementasi kebijakan serta konsekuensinya.

2.2.3 Klasifikasi

Klasifikasi data merupakan suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan objek dalam sebuah basis data dan mengklasifikasikannya ke dalam kelas-kelas dari model (Ente, Sri Astutik Thamrin, Hedi Kuswanto, Samsul Arifin, & Andreza, 2020) pembagian struktur mengenai pada penetapan. Tujuan dari klasifikasi ialah menemukan model dari data latih yang akan membedakan atribut kedalam kategori (label) yang sesuai dengan model (Najib, Nurcahyono, & Setiawan, 2019)

2.2.4 Data Mining

Data mining adalah proses pengumpulan informasi dari data yang besar menggunakan metode statistic, matematika hingga memanfaatkan teknologi *artificial intelligence*. Adapun pengertian lain berasal data mining yaitu teknologi baru yang sangat bermanfaat buat membantu perusahaan menemukan berita yang sangat penting dari gudang data. Beberapa perangkat lunak data mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi pada situasi baru dari data yang mendeskripsikan apa yang terjadi pada masa kemudian (Fatmawati, 2016) data mining juga bagian dari *Knowledge Discovery in Datasbase* (KDD) yang artinya proses ekstraksi informasi yang berguna, tidak di ketahui sebelumnya dan tersembunyi dari data secara garis besar Knowlwdge Discovery in Database (KDD) bisa di jelaskan sebagai berikut (Fatmawati, 2016).

1. Data selection

Data Selection yaitu pemilihan (seleksi) data dari kumpulan data operasional perlu di lakukan sebelum tahap penggalian informasi dalam KDD di mulai. Data hasil seleksi yang akan di gunakan untuk proses data mining di simpan dalam suatu berkas terpisah dari basis data operasional.

2. Pre-Processing/Cleaning

Pre-Processing/Cleaning yaitu proses cleaning diantaranya membuang duplikat data, mengusut data yang inkonsisten serta memperbaiki kesalahan pada data. pada proses dapat di lakukan pula enrichment ialah proses memperbanyak data yang telah di dapat menggunakan data atau informasi lain yg relevan yang pada perlukan buat KDD.

3. Transformation Coding

Transformation Coding ialah proses transformasi dari data yang sudah di pilih, kemudian data yang sudah di pilih tadi di sinkron dalam proses pembuatan data mining.

4. Data mining

Data Mining adalah proses mencari informasi yang menarik pada data terpilih dengan menggunakan teknik atau metode tertentu.

5. Interpretation/Evaluation

Interpretation/Evaluation adalah informasi yang di hasilkan dari proses data mining yang di artikan dengan bentuk lebih mudah dimengerti oleh pihak yang berkepentingan.

2.2.5 Decision Tree c4.5

Pohon keputusan ialah sala satu metode klasifikasi serta prediksi yang sangat kuat dan terkenal dalam penerapan data mining intinya *Decision Tree* mengganti data menjadi pohon keputusan (*Decision tree*) dan aturan-aturan keputusan (rule) pohon keputusan pula bermanfaat untuk mengeksplorasi data menentukan hubungan tersembunyi antara sejumlah calon variable input menggunakan sebuah variable sasaran. Sebuah pohon keputusan mungkin di bangun dengan akurat secara manual atau dapat tumbuh secara otomatis menggunakan menerapkan sala satu atau beberapa algoritma pohon keputusan di buat untuk memodelkan himpunan data yang belum terklasifikasi. banyak algoritma yang di pakai dalam pembentukan pohon keputusan diantaranya ID3, CART dan C4.5 (Fatmawati, 2016) Variabel tujuan umumnya pada kelompokkan dengan pasti serta model pohon keputusan lebih mengarah pada perhitungan probabilitas dari tiap-tiap record kategori tersebut atau untuk mengklasifikasikan record dengan mengelompokkan pada satu kelas. *Decision tree* membuat set rule yg paling efisien dan kemungkinan terkecil yang membuatnya sebagai predictive contoh yang baik. Jika terdapat overlap pada antara dua predictor maka yang terbaik dari keduanya akan di ambil dari system ruleinductin sala satu akan menjadi lemah serta kurang akurat, kelebihan-kelebihan *decision tree* merupakan sebagai berikut :

1. Menyediakan visual result.
2. Dibangun berdasarkan rule-rule yang dapat di pahami dan di mengerti.
3. Bersifat predictive
4. memungkinkan untuk melakuakn prediksi
5. Menampilkan apa yang penting. Algoritma akan mengidentifikasi suatu ser rule tang akan memberikan presentase kemungkinan akan terjadi hal demikian di kemudian hari.

di buat dengan menggunakan sebuah teknik yang di sebut recursive partitioning. Algoritma akan mendefinisikan atribut yang paling relevan dan akan men-split data yang akan ada berdasarkan atribut tersebut setiap partision di sebut sebagai rule. Proses akan di ulang terus untuk setiap subgroup sampai di temukan sebuah good stopping point. Informasi gain measure di gunakan untuk memiliki atribut mana yang akan di tes pada node tree (Zubaidah, Sistem Informasi Manajemen Informatika, 2017)atribut dengan informasi gain yang paling tinggi akan di pilih sebagai test attribute untuk current node (Fatmawati, 2016).

2.2.6 Algoritma C4.5

Konsep dasar dari *algoritma C4.5* yaitu mengubah data menjadi pohon keputusan dan aturan – aturan keputusan (rule). *Algoritma C4.5* menetapkan atribut menjadi kelas yang dapat di terapkan untuk klasifikasi baru (Novandya & Oktra @017). Kelebihan dari *algoritma C4.5* yaitu mudah di pahami , fleksibel dan menarik karena mampu divisualisasikan dalam gambar. Secara umum langkah-langkah *algoritma C4.5* untuk membangun pohon keputusan. adalah sebagai berikut (Ente, Sri Astutik Thamrin, Hedi Kuswanto, Samsul Arifin, & Andreza, 2020).

1. Memilih atribut sebagai akar deng an berdasarkan nilai gain tertinggi pada masing-masing atribut yaitu dengan cara:

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

Keterangan =

S : himpunan kasus

A : atribut

N : jumlah partisi atribut A

$|I|$: jumlah kasus pada partisi ke-I

$|S|$: jumlah kasus pada s

menghitung nilai entropy dapat digunakan rumus;

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

Keterangan =

S : himpunan kasus

A : Atribut

n : jumlah partisi S

P_i : proporsi dari S_i terhadap S.

2. Membuat cabang pada masing-masing nilai
3. Membagi kasus dalam cabang
4. Mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

2.2.7 Python

Python adalah sebuah bahasa pemrograman yang cukup terkenal yang memiliki tujuan untuk mendukung pemrograman yang berorientasi objek dan dapat berjalan di berbagai macam platform system operasi seperti PCs, Macintosh, UNIX beberapa kelebihan dari bahasa pemrograman python (Ginting, Kusriani, & Emha Taufiq, 2020) di antaran lain :

1. Pengembangan program di lakukan dengan cepat dan coding yang lebih sedikit.
2. Mendukung multi platfrom
3. Python mudah di pelajari
4. Memiliki system pengolahan memori yang otomatis.
5. Python bersifat Object Oriented programming.

Confusion Matrix adalah sebuah metode yang dapat di gunakan untuk melakukan perhitungan akurasi pada konsep dalam facts Mining. Evaluasi dengan menggunakan metode confusion matrix menghasilkan nilai akurasi, presisi dan recall. Akurasi dalam klasifikasi dalam records mining merupakan persentase ketepatan pada file statistics yang telah di klasifikasikan secara tepat dan di lakukan pengujian pada hasil klasifikasi (Ginting, Kusri, & Emha Taufiq, 2020) presisi atau self belief merupakan proporsi pada kasus yang diprediksi positif, dimana statistics yang sebenarnya juga positif. recall atau sensitivity adalah proporsi kasus yang di prediksi positif dimana data yang sebenarnya juga positif. don't forget atau sensitivity adalah proporsi kasus positif yang di prediksi dengan benar (Abdillah, 2011)

