

Penerapan Algoritma Iterative Dichotomiser 3 (ID3) Untuk Klasifikasi Penyakit Tifoid

Muhammad Yogi Firmansyah NIM 1510651166

email: myogifirmansyah@gmail.com

Pembimbing : Dewi Lusiana, Ir.MT

Prodi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember

Abstrak – Algoritma ID3 adalah algoritma data mining dengan mencari hasil klasifikasi menggunakan sebuah pohon keputusan pada penelitian ini meneliti data pasien yang terkena penyakit tifoid menggunakan sebuah pohon keputusan

Tujuan dari penelitian ini adalah menerapkan algoritma id3 untuk mengetahui hasil klasifikasi data pasien yang terdampak penyakit tifoid pada puskesmas sukorambi jember

Hasil dari penelitian ini adalah diperoleh nilai akurasi sebesar 89,44% dan nilai presisi sebesar 89,48% data diuji menggunakan beberapa fold dengan 5 skenario uji

Tujuan penelitian adalah mengetahui nilai akurasi dan presisi menggunakan metode Decision Tree Iterative Dichotomiser 3 (ID3).

Hasil dari penelitian ini ialah mendapatkan nilai akurasi dan presisi dengan metode klasifikasi data mining algoritma ID3 sehingga dapat memprediksi data pasien yang terdampak penyakit tifoid pada puskesmas sukorambi jember

Kata kunci : Algoritma ID3, Tifoid, Puskesmas sukorambi jember

BAB I PENDAHULUAN

Demam Tifoid merupakan infeksi akut yang disebabkan oleh bakteri *Salmonella enterica* reservoir typhi, umumnya disebut *salmonella typhi* (*S.typhi*), jumlah kasus demam Tifoid di seluruh dunia diperkirakan terdapat 21 juta kasus dengan 128.000 sampai 161.000 kematian setiap tahun, kasus terbanyak terdapat di asia selatan dan asia tenggara (Anggit, 2018).

Pada puskesmas khususnya di daerah sukorambi jember demam Tifoid masih relative tinggi pada tahun 2019. masih jarang ditemui data layanan informasi yang terinci pada puskesmas sehingga para pasien masih bingung pada gejala-gejala apa saja yang dialami pada penyakit yang diderita. sehingga dibutuhkan suatu informasi-informasi tentang penyakit kepada para pasiennya agar pasien dapat mengetahui apa saja ciri pada sebuah penyakit. disini layanan informasi yang cocok dan mudah untuk dipahami bagi pasien adalah membentuk sebuah bagan atau sebuah tabel pohon keputusan. dalam membentuk pohon keputusan dibutuhkan sebuah metode Klasifikasi.

Klasifikasi merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. model itu sendiri bisa berupa aturan “jika-maka”, berupa decision tree formula matematis atau neural network. Metode-metode klasifikasi antara lain C4.5, RainForest, Naïve Bayes, Neural Network, ID3, case-based reasoning, dan K-Nearest Neighbor (Arriawati A S, 2011).

Dari beberapa metode klasifikasi tersebut terdapat metode yang memiliki tingkat akurasi tinggi yaitu Decision Tree. Penelitian yang dilakukan oleh Gaussian (2015) yang berjudul “ Algoritma Iterative Dichotomiser 3 (ID3) untuk mengidentifikasi data rekam medis. pada penelitian ini peneliti memakai 84 data sampel. diperoleh akurasi sebesar 90% menggunakan confusion matrix.

Dari penelitian tersebut menunjukkan bahwa algoritma Decision Tree ID3 memiliki tingkat akurasi yang tinggi pada proses klasifikasi data. metode ini tentu bisa diterapkan pada data pasien yang terdampak Penyakit Tifoid. sehingga tujuan dari diterapkannya metode ini untuk memudahkan puskesmas dalam menangani pasien yang terkena penyakit Tifoid.

Algoritma Iterative Dichotomiser 3 (ID3) merupakan salah metode dalam data mining. Data Mining mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang akademi, bisnis hingga medis. ID3 adalah algoritma decision tree learning (algoritma pembelajaran pohon) yang paling dasar. Algoritma ini melakukan pencarian secara menyeluruh pada semua kemungkinan pohon keputusan. Pembentukan pohon klasifikasi dengan algoritma ID3 melalui dua langkah, yaitu menghitung nilai entropy dan menghitung nilai information gain dari setiap variabel. ID3 dapat menyelesaikan kasus pada berbagai bidang salah satunya dapat diterapkan pada bidang kesehatan (Santosa, 2007).

Oleh karena itu peneliti melakukan sebuah penelitian dengan mengimplementasi data dengan pengambilan data training 20 dan data testing 20 menggunakan data yang peneliti ambil dari Puskesmas Sukorambi Jember, dengan kriteria masing-masing data yang berbeda meliputi variable, tipe data dan jumlah data, Hal ini bertujuan untuk menganalisa hasil akurasi. Berdasarkan pembahasan diatas peneliti mengambil judul “Penerapan Algoritma Decision Tree (ID3) Untuk Klasifikasi Penyakit Tifoid”.

1.2 Rumusan masalah

Berdasarkan permasalahan-permasalahan yang ada pada latar belakang penelitian ini, maka dapat diidentifikasi permasalahannya adalah Berapa akurasi dan presisi yang dihasilkan menggunakan metode Decision Tree Iterative Dechotomiser 3 (ID3) ?

Tujuan

Sesuai dengan masalah yang dirumuskan, maka tujuan dari penelitian ini adalah mengetahui nilai akurasi dan presisi menggunakan metode Decision Tree Iterative Dechotomiser 3 (ID3).

Manfaat

Berdasarkan Tujuan Penelitian yang hendak dicapai, Maka penelitian ini diharapkan memiliki manfaat bagi puskesmas khususnya daerah sukorambi jember baik secara langsung maupun tidak langsung adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Bagi instansi kesehatan

Membantu instansi puskesmas dalam memberikan informasi pada para pasien yang terdampak penyakit Tifoid.

2. Bagi Pasien

Membantu pasien untuk lebih mengenali ciri-ciri pada penyakit tipes dan gejalanya.

3. Bagi Peneliti

Manfaat yang didapat bagi peneliti adalah mampu mengimplementasikan ilmu yang didapat dari perkuliahan sehingga dapat diterapkan pada dunia nyata.

Batasan masalah

Agar permasalahan lebih terarah dan tidak menyimpang dari pokok permasalahan, maka dibutuhkan batasan masalah sebagai Berikut :

1. Ruang lingkup penelitian saat ini hanya dibatasi hanya di daerah kabupaten jember khususnya puskesmas Sukorambi.

2. Data yang digunakan pada penelitian ini adalah data dari Puskesmas Sukorambi yang berjumlah 210 data.

3. Output yang digunakan ada 2 yaitu Tifoid Fever dan Tifoid Gea.

BAB. II LANDASAN TEORI

Demam Tifoid atau thypus abdominalis adalah penyakit infeksi yang biasanya mengenai saluran pencernaan dengan gejala demam lebih dari 7 hari, gangguan pada saluran cernadan gangguan kesadaran. dalam masyarakat penyakit ini dikenal dengan nama Tipes atau thypus (Sudoyo 2009).

Penyakit ini disebabkan oleh salmonella typhosa dan hanya didapatkan pada manusia. Penularan penyakit ini hamper selalu terjadi melalui makanan dan minuman yang terkontaminasi. Kata epidemiologi berasal dari bahasa yunani, epi berarti pada/tentang, demos berarti penduduk, dan logos berarti ilmu.

Epidemiologi adalah ilmu yang mempelajari berbagai faktor dan kondisi yang mempengaruhi suatu kejadian dan penyebaran keadaan sehat, sakit, kerusakan jaringan, kelumpuhan serta kematian (Rajab, 2009).

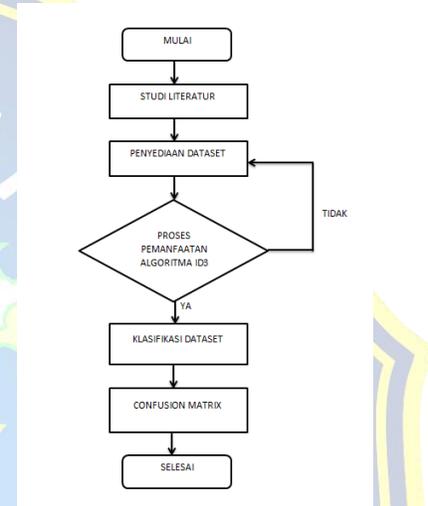
Pada Demam Tifoid yang umum terjadi pada masyarakat adalah demam Tifoid Fever dan Tifoid Gea. Tifoid

Fever merupakan demam yang disertai dengan suhu tubuh yang tinggi sedangkan Tifoid Gea adalah demam tifoid yang disertai dengan Gangguan saluran pencernaan.

BAB. III METODOLOGI PENELITIAN

Tahapan Penelitian

Adapun metodologi penelitian yang dilakukan dalam pemanfaatan algoritma Iterative Dichotomiser 3 (ID3) untuk mengidentifikasi pasien yang terdeteksi penyakit Tifoid, memiliki langkah-langkah yang dapat digambarkan sebagai berikut:



Gambar 1. Flowchart Metode Penelitian

Studi Literatur

Tahap study literature terkait permasalahan, wawancara secara langsung terhadap sumbernya dan pengambilan data pasien yang terdampak penyakit tifoid pada puskesmas sukorambi Jember. Study literatur dilakukan dengan cara mencari bahan materi yang berhubungan dengan permasalahan, mempelajari teori – teori mengenai pengetahuan tentang metode Decisin Tree Iterative Dechotomiser 3 (ID3). Pencarian materi melalui sumber dari internet, jurnal yang berkaitan dengan penelitian yang dilakukan.

Penyediaan Dataset

Proses penyediaan dataset dilakukan dengan cara meminta dataset yang diperoleh dari Puskesmas Sukorambi Jember. Dari dataset tersebut kemudian di olah dengan menggunakan Data Mining dengan menggunakan metode dari ID3. Total jumlah data 210 pada dataset tersebut memiliki 5 atribut yaitu : Sakit Perut, Ruam Kulit, Radang Tenggorokan, Mual, Sakit Kepala.

Pemanfaatan Algoritma ID3

Dalam pemanfaatan algoritma ID3 peneliti menggunakan data training sebanyak 20 data sample. tahapan pertama kita mencari Simpul Root atau titik awal dari Decision Tree kemudian tahap selanjutnya sitentukan simpul perantara setiap simpul perantara berhubungan dengan suatu pertanyaan atau pengujian, pohon yang terakhir adalah

simpul leaf yaitu keputusan akhir atau kelas target untuk suatu pohon keputusan.

Untuk menentukan ketiga tahapan tersebut kita melakukan perhitungan terlebih dahulu dengan mencari nilai entropi dan gain pada data training yang tersedia untuk memudahkan perhitungan kita bisa menggunakan Microsoft excel dan membentuk table agar memudahkan kita dalam menentukan entropi dan gain yang terbesar hingga yang terkecil.

Standart Pengklasifikasian Data

Proses Training

Pada proses training yaitu memasukan data sampel kedalam tabel yang disiapkan untuk proses perhitungan. Tabel tersebut meliputi atribut, jumlah data yang sudah terklasifikasi berdasarkan target yang sudah ditentukan, dalam kasus ini yaitu output yang dihasilkan berupa Tifoid Fever (Demam) atau Tifoid GEA (Pencernaan), serta kolom nilai entropi dan gain. Tahapan selanjutnya adalah penerapan algoritma ID3 yaitu menghitung nilai Entropy dan Gain pada tiap-tiap atribut untuk dijadikan bentuk Tree. Tree merupakan bentuk aturan klasifikasi yang akan diterapkan pada proses testing, (Mashlahah, 2013).

Proses Testing

Pada proses testing ini langkah yang dilakukan yaitu memasukan data uji atau data prediksi. Atribut yang digunakan pada proses testing harus sesuai dengan atribut pada proses training. Setiap data atribut akan dibandingkan dengan aturan yang sudah terbentuk pada perhitungan data training sebelumnya. Selanjutnya data tersebut akan diklasifikasikan berdasarkan target yang ingin diketahui yaitu data pasien yang mengalami penyakit Tifoid fever (Demam) atau Tifoid GEA (Pencernaan), (Mashlahah, 2013).

Tabel 3.1 Data Training

| No | Nama Pasien | Sakit Perut | Radang Tenggorokan | Ruam Kulit | Mual | Sakit Kepala | Output |
|----|-----------------|-------------|--------------------|------------|-------|--------------|--------------|
| 1 | Vifi | Tidak | Tidak | Ya | Ya | Ya | Typoid Fever |
| 2 | Putra Ramadhani | Ya | Ya | Tidak | Ya | Tidak | Typoid GEA |
| 3 | Zammil | Tidak | Tidak | Ya | Tidak | Tidak | Typoid Fever |
| 4 | Sayutik | Ya | Tidak | Tidak | Ya | Ya | Typoid GEA |
| 5 | Rwanda | Tidak | Ya | Tidak | Tidak | Ya | Typoid Fever |
| 6 | Ernawati | Tidak | Tidak | Ya | Tidak | Ya | Typoid Fever |
| 7 | Nurul Islamiyah | Tidak | Ya | Tidak | Tidak | Tidak | Typoid GEA |
| 8 | M.Nur | Ya | Tidak | Tidak | Ya | Tidak | Typoid GEA |
| 9 | Masut | Ya | Tidak | Ya | Ya | Tidak | Typoid Fever |
| 10 | Azzahra | Tidak | Tidak | Ya | Tidak | Ya | Typoid Fever |
| 11 | Ahmad Wahyu | Tidak | Tidak | Tidak | Ya | Tidak | Typoid Fever |
| 12 | Moh.Bintang | Ya | Tidak | Ya | Tidak | Ya | Typoid Fever |
| 13 | Winda | Tidak | Ya | Tidak | Ya | Tidak | Typoid Fever |
| 14 | Nabila | Ya | Tidak | Ya | Tidak | Ya | Typoid Fever |
| 15 | Mutama | Ya | Ya | Tidak | Tidak | Ya | Typoid GEA |
| 16 | Hotimah | Ya | Ya | Tidak | Ya | Tidak | Typoid GEA |
| 17 | Sonia.P | Tidak | Ya | Ya | Ya | Tidak | Typoid Fever |
| 18 | Asharfa.A | Tidak | Tidak | Ya | Tidak | Ya | Typoid GEA |
| 19 | Rahman | Ya | Ya | Tidak | Ya | Tidak | Typoid GEA |
| 20 | Gufron | Tidak | Tidak | Tidak | Ya | Tidak | Typoid GEA |

Dari data Training diatas dihitung nilai Entropy dan Gain untuk masing-masing atribut.

- Atribut : Sakit Perut,Ruam Kulit,Radang Tenggorokan,Mual,Sakit Kepala.
- Output : Output (Typoid Fever dan Typoid GEA).
- Hitung Entropy dan Gain masing-masing atribut.

$$\text{Entropy (Total)} = (- 10/20 \text{ [log]}_2 (10/20)) + (- 10/20 \text{ [log]}_2 (10/20)) = 1$$

$$\text{Entropy (Sakit Perut,Ya)} = (- 3/9 \text{ [log]}_2 (3/9)) + (- 6/9 \text{ [log]}_2 (6/9)) = 0,918295834$$

$$\text{Entropy (Sakit Perut,Tidak)} = (- 7/11 \text{ [log]}_2 (7/11)) + (- 4/11 \text{ [log]}_2 (4/11)) = 0,945660305$$

$$\text{Gain (Total,Sakit Perut)} = 1 - ((9/20 *0,918295834) + (11/20*0,945660305)) = 1,933346293$$

$$\text{Entropy (Radang Tenggorokan,Ya)} = (- 3/10 \text{ [log]}_2 (3/10)) + (- 7/10 \text{ [log]}_2 (7/10)) = 0,881290899$$

$$\text{Entropy (Radang Tenggorokan,Tidak)} = (- 7/10 \text{ [log]}_2 (7/10)) + (- 3/10 \text{ [log]}_2 (3/10)) = 0,881290899$$

$$\text{Gain (Total,Radang Tenggorokan)} = 1 - ((10/20 *0,881290899) + (10/20*0,881290899)) = 1,881290899$$

$$\text{Entropy (Ruam Kulit,Ya)} = (- 7/9 \text{ [log]}_2 (7/9)) + (- 2/9 \text{ [log]}_2 (2/9)) = 0,764204507$$

$$\text{Entropy (Ruam Kulit,Tidak)} = (- 3/11 \text{ [log]}_2 (3/11)) + (- 8/11 \text{ [log]}_2 (8/11)) = 0,845350937$$

$$\text{Gain (Total,Ruam Kulit)} = 1 - ((9/20 *0,764204507) + (11/20*0,845350937)) = 1,808835043$$

$$\text{Entropy (Mual,Ya)} = (- 4/11 \text{ [log]}_2 (4/11)) + (- 7/11 \text{ [log]}_2 (7/11)) = 0,945660305$$

$$\text{Entropy (Mual,Tidak)} = (- 6/9 \text{ [log]}_2 (6/9)) + (- 3/9 \text{ [log]}_2 (3/9)) = 0,918295834$$

$$\text{Gain (Total,Mual)} = 1 - ((11/20 *0,945660305) + (9/20*0,918295834)) = 1,933346293$$

$$\text{Entropy (Sakit Kepala,Ya)} = (- 6/10 \text{ [log]}_2 (6/10)) + (- 4/10 \text{ [log]}_2 (4/10)) = 0,945660305$$

$$\text{Entropy (Sakit Kepala,Tidak)} = (- 4/10 \text{ [log]}_2 (4/10)) + (- 6/10 \text{ [log]}_2 (6/10)) = 0,918295834$$

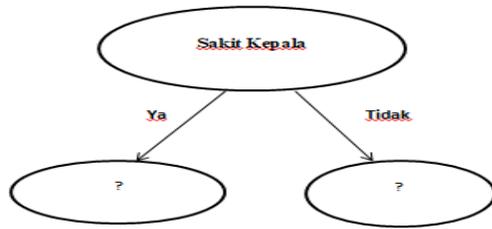
$$\text{Gain (Total,Sakit Kepala)} = 1 - ((10/20 *0,945660305) + (10/20*0,918295834)) = 1,970950594$$

Tabel 3.2 Mencari Node Akar Pertama Dengan Gain Terbesar

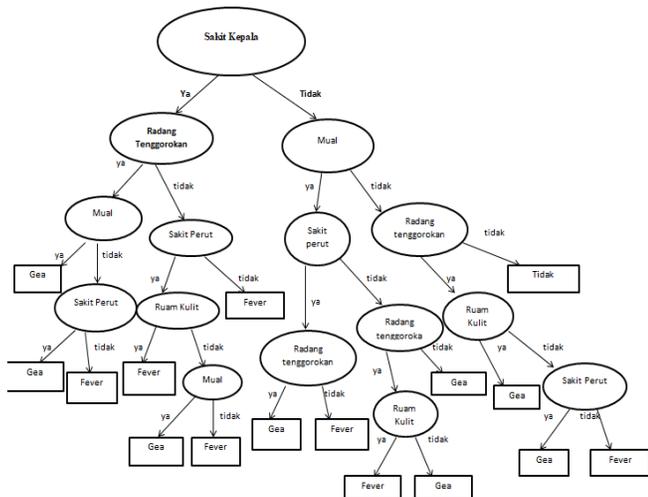
| NO | ATRIBUT | JUMLAH KASUS | TIFOID FEVER(S1) | TIFOID GEA(S2) | ENTROPY | GAIN |
|----|--------------------|--------------|------------------|----------------|---------|-------------|
| 1 | TOTAL | 20 | 10 | 10 | 1 | |
| 2 | Sakit Perut | | | | | 1,933346293 |
| | | Ya | 9 | 3 | 6 | 0,918295834 |
| | | Tidak | 11 | 7 | 4 | 0,945660305 |
| 3 | Radang Tenggorokan | | | | | 1,881290899 |
| | | Ya | 10 | 3 | 7 | 0,881290899 |
| | | Tidak | 10 | 7 | 3 | 0,881290899 |
| 4 | Ruam Kulit | | | | | 1,808835043 |
| | | Ya | 9 | 7 | 2 | 0,764204507 |
| | | Tidak | 11 | 3 | 8 | 0,845350937 |
| 5 | Mual | | | | | 1,933346293 |
| | | Ya | 11 | 4 | 7 | 0,945660305 |
| | | Tidak | 9 | 6 | 3 | 0,918295834 |
| 6 | Sakit Kepala | | | | | 1,970950594 |
| | | Ya | 10 | 6 | 4 | 0,970950594 |
| | | Tidak | 10 | 4 | 6 | 0,970950594 |

Pada tabel 1 diatas dipilih gain terbesar, dan gain terbesar terdapat pada atribut Sakit Kepala.

- Sekarang kita mendapatkan rule yang pertama yaitu atribut Sakit Kepala sebagai prioritas.Dan untuk Sakit Kepala Ya dan Tidak masih harus dicari karena belum bernilai 0.Maka dari itu akan dihitung kembali untuk atribut Sakit Kepala.
- Langkah pertama dalam membangun pohon keputusan yaitu memilih atribut sebagai akar. Dari hasil perhitungan diatas dapat digambarkan pohon keputusan sementara.



Gambar 3.2 Hasil Pohon Keputusan Node ke 1 Setelah dilakukan perhitungan dengan mengutamakan node yang terbesar hingga yang terkecil diperoleh sebuah pohon keputusan seperti gambar dibawah ini



Gambar 3.14 keputusan Node ke 13

- Pohon keputusan telah selesai karna semua cabang telah diketahui Hasilnya
- Selanjutnya dari hasil pohon keputusan tersebut kita bisa membuat Tabel Hasil Klasifikasinya
- Tabel Klasifikasi digunakan untuk menghitung Confusion Matrix pada data sehingga memperoleh nilai akurasi dan presisinya.

3.6 Pengukuran Ketepatan Hasil Klasifikasi Menggunakan Data Testing

Setelah didapatkan secara utuh hasil klasifikasi algoritma ID3 berupa pohon keputusan, langkah selanjutnya adalah mengukur tingkat akurasi dan presisi menggunakan data testing. Ukuran sampel pengujian sebanyak 20 kasus.

hasil klasifikasi menggunakan data testing

| No | Nama Pasien | Jenis Kelamin | Umur | radang tenggorok | muall | Sakit Kepala | Output | hasil klasifikasi | kriteria |
|----|-----------------|---------------|-----------|------------------|-------|--------------|--------------|-------------------|----------|
| 1 | Vifi | Perempuan | Anak-Anak | Tidak | Ya | Ya | Typoid Fever | Typoid Fever | TP |
| 2 | Putra Ramadhani | Laki-Laki | Anak-Anak | Ya | Tidak | Ya | Typoid GEA | Typoid GEA | TN |
| 3 | Zammil | Laki-Laki | Remaja | Ya | Tidak | Tidak | Typoid Fever | Typoid Fever | TP |
| 4 | Sayutik | Perempuan | Remaja | Tidak | Tidak | Ya | Typoid GEA | Typoid GEA | TN |
| 5 | Rwanda | Perempuan | Anak-Anak | Ya | Tidak | Ya | Typoid Fever | Typoid Fever | TP |
| 6 | Ernawati | Perempuan | Anak-Anak | Tidak | Tidak | Ya | Typoid Fever | Typoid Fever | TP |
| 7 | Nurul Islamiyah | Perempuan | Remaja | Ya | Tidak | Tidak | Typoid GEA | Typoid GEA | TN |
| 8 | M.Nur | Laki-Laki | Remaja | Ya | Ya | Tidak | Typoid GEA | Typoid Fever | FN |
| 9 | Masut | Laki-Laki | Remaja | Tidak | Ya | Tidak | Typoid Fever | Typoid GEA | FP |
| 10 | Azzahra | Perempuan | Anak-Anak | Tidak | Ya | Ya | Typoid Fever | Typoid GEA | FP |
| 11 | Ahmad Wahyu | Laki-Laki | Remaja | Tidak | Ya | Tidak | Typoid Fever | Typoid Fever | TP |
| 12 | Moh.Bintang | Laki-Laki | Anak-Anak | Tidak | Ya | Tidak | Typoid Fever | Typoid Fever | TP |
| 13 | winda | Perempuan | Remaja | Ya | Ya | Tidak | Typoid Fever | Typoid Fever | TP |
| 14 | Nabila | Perempuan | Anak-Anak | Tidak | Tidak | Ya | Typoid Fever | Typoid Fever | TP |
| 15 | Mutama | Perempuan | Remaja | Ya | Tidak | Ya | Typoid GEA | Typoid Fever | FN |
| 16 | Hotimah | Perempuan | Anak-Anak | Ya | Ya | Tidak | Typoid GEA | Typoid Fever | FN |
| 17 | Sonia.P | Laki-Laki | Anak-Anak | Ya | Ya | Tidak | Typoid Fever | Typoid Fever | TP |
| 18 | Asharla.A | Perempuan | Remaja | Tidak | Tidak | Ya | Typoid GEA | Typoid Fever | FN |
| 19 | Rahman | Laki-Laki | Remaja | Tidak | Tidak | Ya | Typoid GEA | Typoid GEA | TP |
| 20 | gufon | Laki-Laki | Anak-Anak | Tidak | Ya | Tidak | Typoid GEA | Typoid GEA | TP |

3.6 Confusion Matrix

Dengan ketentuan jika klasifikasi bernilai "Fever" dan data asli bernilai "Fever" maka kriteria true positive (TP), jika klasifikasi bernilai "Fever" dan data asli bernilai "Gea" kriteria false positive (FP), jika klasifikasi bernilai "Gea" dan data asli bernilai "Fever" kriteria false negative (FN), jika klasifikasi bernilai "Gea" dan data asli bernilai "Gea" maka kriteria true negative (TN).

Ketepatan klasifikasi yang telah dihitung sebagai berikut :

Tabel 3.16 Hasil Accuracy Precision

| Kriteria | Jumlah |
|-----------|--------|
| TP | 11 |
| TN | 3 |
| FP | 2 |
| FN | 4 |
| Accuracy | 70% |
| Precision | 84% |

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100$$

$$= \frac{(11+3)}{(11+3+2+4)} \times 100$$

$$= \frac{14}{20} \times 100$$

$$= 70\%$$

$$Precision = \frac{TP}{(TP+FP)} \times 100$$

$$= \frac{11}{(11+2)} \times 100$$

$$= \frac{11}{13} \times 100$$

$$= 84\%$$

Berdasarkan hasil tingkat akurasi dan presisi dari data Testing diperoleh tingkat akurasi sebesar 70 % dan presisi 84%. sehingga algoritma ID3 cukup baik digunakan untuk memprediksi pada data pasien selanjutnya.

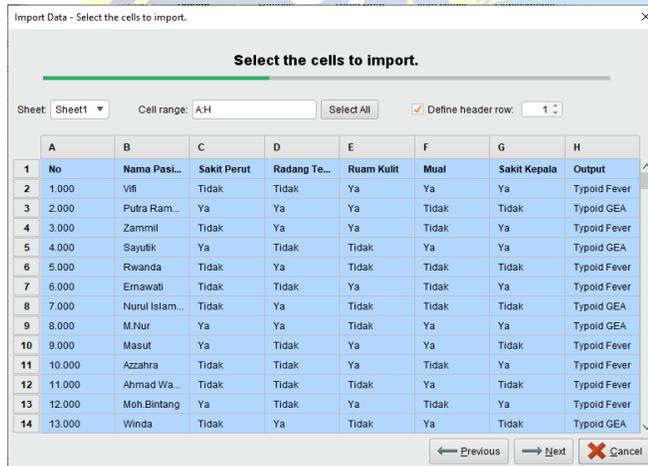
BAB. IV HASIL DAN PEMBAHASAN

4.1 Gambaran Dataset

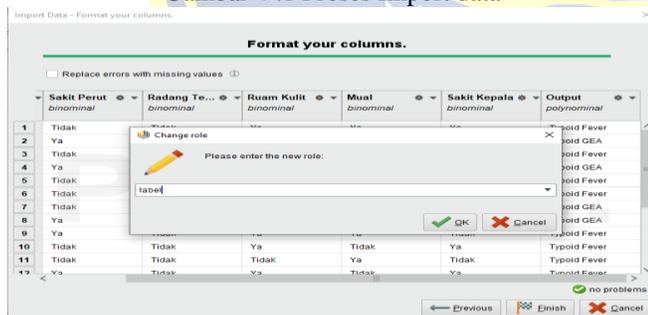
Data penelitian yang digunakan diambil dari Puskesmas Sukorambi Jember. Data set tersebut diminta dari petugas yang menangani tentang Penyakit Tifoid, pada data set tersebut memiliki 5 (lima) atribut dan satu output yang terdiri dari dua kelas, yaitu “TIFOID FEVER” Yakni Tifoid yang disertai dengan demam yang tinggi dan “TIFOID GEA”Yakni Tifoid yang disertai dengan gangguan saluran pencernaan. Berikut adalah gambaran atribut dalam dataset. Jumlah pasien yang terdaftar dalam data set berjumlah 210 orang. Untuk keterangan atribut di jelaskan pada tabel 4.1 di bawah ini.

Implementasi Dalam Rapid Miner

Data uji dan data latih yang sudah di bagi tadi akan dimasukkan ke dalam Rapid Miner untuk di proses, data yang pertama di masukan adalah percobaan pada skenario yang telah dibagi menjadi 2,3,5,7 dan 10 berikut merupakan pengolahan data dengan menggunakan algoritma ID3 pada Rapid Miner. Langkah-langkah dalam implementasi Rapid Miner untuk Percobaan di jelaskan pada gambar-gambar di bawah ini.

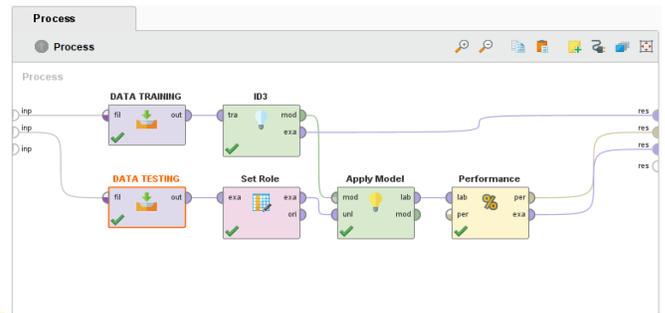


Gambar 4.1 Proses Import data



Gambar 4.2 Proses penentuan label pada Output

Proses inputkan data yang pertama kita import data pada excel kedalam rapidminer sebelum di finish kita ubah dulu role pada data untuk atribut yang pertama polynomial kita ubah ke binomial dan untuk output kita ubah menjadi label



Gambar 4.3 Proses penerapan algoritma ID3 memasukkan data training dan data testing selanjutnya inputkan algoritma id3 untuk bagian data testing tambahkan set role berikut adalah setingan set role pada proses id3 atribut name diubah menjadi output dan target role diubah label Hasil

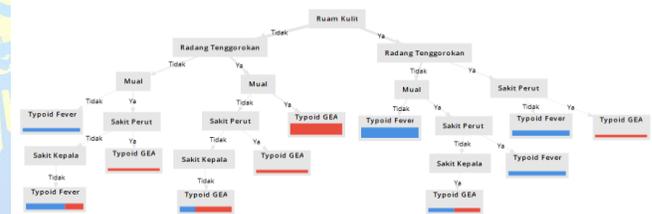
Dari Hasil pada pengujian k-fold 5 skenario 2,3,5,7 dan 10 berikut adalah tampilan semua akurasi dan presisi yang didapatkan :

| Skenario | 2 Fold | | 3 Fold | | | 5 Fold | | | | |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | P-1 | P-2 | P-1 | P-2 | P-3 | P-1 | P-2 | P-3 | P-4 | P-5 |
| Akurasi | 87.62% | 84.76% | 86.43% | 86.43% | 86.43% | 86.90% | 86.31% | 86.31% | 88.69% | 77.98% |
| Presisi | 91.21% | 85.24% | 90% | 88.27% | 87.08% | 90.35% | 87.46% | 88.02% | 89.24% | 80.33% |

| Skenario | 7 Fold | | | | | | | |
|----------|--------|--------|--------|--------|---------------|---------------|--------|--------|
| | P-1 | P-2 | P-3 | P-4 | P-5 | P-6 | P-7 | |
| Akurasi | 86.67% | 86.67% | 86.67% | 86.67% | 71.67% | 89.44% | 80% | 82.86% |
| Presisi | 90.08% | 89.54% | 90.08% | 80.23% | 89.48% | 82.5% | 84.94% | |

| Skenario | 10 Fold | | | | | | | | | |
|----------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | P-1 | P-2 | P-3 | P-4 | P-5 | P-6 | P-7 | P-8 | P-9 | P-10 |
| Akurasi | 86.77% | 86.77% | 86.77% | 86.71% | 86.77% | 86.24% | 85.19% | 81.48% | 80.95% | 80.95% |
| Presisi | 90.15% | 87.7% | 87.68% | 87.7% | 87.7% | 88.29% | 86.43% | 87.03% | 83.29% | 83.26% |

Dari hasil kesimpulan yang diperoleh akurasi dan Presisi tertingginya terdapat pada 7 Fold Skenario P-5 yakni Akurasi : 89.44%
Presisi : 89.48%



Keterangan : = Typoid Fever = Typoid GEA

BAB V. PENUTUP

Kesimpulan

Setelah melakukan analisis dan implementasi beserta pengujian dalam pemanfaatan algoritma ID3 untuk klasifikasi penyakit tifoid, maka didapatkan kesimpulan bahwa:

1. Algoritma ID3 dapat menghasilkan tingkat akurasi dan presisi yang konsisten dengan melakukan pengujian sebanyak 5 fold dengan skenario 2,3,5,7 dan 10 dengan data uji yang berbeda-beda, yaitu mendapatkan akurasi tertinggi pada semua skenario sebesar 89,44% sedangkan presisi tertingginya mendapatkan nilai sebesar 89,48% pada data 7 fold scenario P-5

2. Dengan demikian algoritma ID3 sangat baik digunakan dalam mendeteksi pasien yang terdampak penyakit tifoid.

Saran

Berdasarkan penelitian yang telah dilakukan, diajukan saran penelitian sebagai berikut:

1. Untuk memperoleh nilai akurasi dan presisi yang tinggi pengujian bisa dilakukan dengan menggunakan beberapa data uji scenario didalamnya.

2. Pengambilan data sample yang baik bisa dilakukan dengan mengambil data bagian atas, tengah, dan bawah pada keseluruhan dataset.

DAFTAR PUSTAKA

Arriawati, A. S. (2011) Klasifikasi Proses untuk Membedakan Konsep atau Kelas Data. *Jurnal Informatika* vol. 8, No.1

Ariadni, R., Arieshanti, I. (2015), Implementasi Metode Pohon Keputusan Untuk Klasifikasi Data Dengan Nilai Fitur Yang Tidak Pasti. *ITS, Surabaya*.

Gaussian. (2015), Algoritma ID3 Untuk Mengidentifikasi Data Rekam Medis. *Jurnal Gaussian*, Vol.4, No.2, Tahun 2015, Halaman 237-246. <http://ejournal-sl.undip.ac.id/index.php/Gaussian>.

H. Sulastris and A.I. Gufroni, (2017), Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. *Jurnal national Teknologi dan Sistem Informasi*, vol.03 No.02, pp.299-305, 2017, doi: <https://doi.org/10.25077/TEKNOSI.v3i2.2017.299-305>.

Hidayatullah, Sarif. (2019) Pemanfaatan Algoritma ID3 untuk Rekomendasi Pemberian Bebas Bersyarat Kepada Narapidana Atas Penyalahgunaan Narkoba. *Universitas Muhammadiyah Jember*.

Prehamukti, Anggit Aprindrian. (2018), Faktor Lingkungan dan Perilaku Terhadap Kejadian Demam Tifoid. *Universitas Negeri Semarang*

Santosa, B. (2007), Teknik Pemanfaatan Data untuk Keperluan Bisnis. *Graha Ilmu: Yogyakarta*

Sudoyo. (2009), Buku Ajar Penyakit Dalam. *Jakarta: Interna Publishing*.

Rajab, Wahyudin. (2009), Buku Ajar Epidemiologi Untuk Mahasiswa Kebidanan. *Jakarta: EGC*.