

3190-8482-2-SM

by jiko unkhair

Submission date: 07-Jun-2021 12:31AM (UTC-0400)

Submission ID: 1601574873

File name: 3190-8482-2-SM.docx (175.97K)

Word count: 3540

Character count: 21026

PENERAPAN NAÏVE BAYES UNTUK KLASIFIKASI TINGKAT RISIKO DIAGNOSIS GIGI DI UPTD PUSKESMAS CINGAMBUL

Elin Nurlia¹, Mohamad Jajuli², Intan Purnamasari³

¹⁵
^{1,2,3}Program Studi Teknik Informatika, ²⁶Jurusan Ilmu Komputer, Universitas Singaperbangsa Karawang
Email: ¹elin.nurlia17088@student.unsika.ac.id, ²mo²mad.jajuli@unsika.ac.id,
³intan.purnama@staff.unsika.ac.id

(Naskah masuk: dd mmm yyyy, diterima untuk diterbitkan: dd mmm yyyy)

Abstrak

Penyakit gigi merupakan kondisi yang dialami ketika muncul rasa sakit pada organ gigi dan mulut karena gigi berlubang atau gangguan lainnya. Kesehatan gigi sangat penting sebab gigi termasuk organ penting pada sistem pencernaan, jika terjadi gangguan dapat menyebabkan masalah kesehatan lainnya. Bahkan peradangan gigi dan gusi yang parah jika tidak segera ditangani dapat menimbulkan penyakit mematikan seperti stroke, diabetes dan jantung. Kasus penyakit gigi yang terjadi di UPTD Puskesmas Cingambul setiap tahunnya masih mengalami peningkatan, bahkan penyakit dengan kondisi yang parah sering dikeluhkan. Namun pelayanan kesehatan gigi di sana masih terbatas karena tidak ada dokter yang bertugas. Oleh karena itu, diperlukan sebuah metode yang mampu mengklasifikasikan tingkat risiko diagnosis gigi agar dapat merencanakan penanganan yang tepat. Penelitian ini bertujuan untuk menerapkan algoritma *Naïve Bayes* dengan metodologi *Knowledge Discovery in Database (KDD)* untuk mengklasifikasikan diagnosis gigi berdasarkan risiko parah, sedang dan ringan. Pengujian penerapan algoritma *Naïve Bayes* dilakukan dengan sembilan skenario *k-fold cross-validation*. Penggunaan *3-fold cross-validation* menunjukkan performa terbaik dengan akurasi 93,33% dan nilai *kappa* 0,855 yang termasuk kategori klasifikasi *Very Good*. Sedangkan rata-rata hasil pengujian memperoleh *accuracy* 92,32% dan nilai *kappa* 0,833 yang termasuk kategori klasifikasi *Very Good*. Sehingga, algoritma *Naïve Bayes* dinilai mampu mengklasifikasikan tingkat risiko diagnosis gigi yang terjadi di UPTD Puskesmas Cingambul dengan baik karena memiliki nilai *accuracy* cukup baik dan nilai *kappa* yang termasuk kategori *Very Good*.

Kata kunci: *cross-validation, diagnosis gigi, KDD, naïve bayes*

IMPLEMENTATION NAÏVE BAYES FOR CLASSIFICATION DENTAL DIAGNOSIS RISK LEVEL IN UPTD CINGAMBUL PUBLIC HEALTH CENTER

Abstract

Dental disease is a condition that is experienced when pain occurs in the dental and oral organs due to cavities or other disorders. Dental health is very important because teeth are an important organ in the digestive system, in case of disruption can cause other health problems. Even severe inflammation of the teeth and gums if not treated immediately can lead to deadly diseases such as stroke, diabetes and heart disease. Cases of dental disease that occur in UPTD Cingambul Public Health Center every year still experience in increase, even diseases with severe conditions are often complained. But dental health services there are still limited because there are no doctors on duty. Therefore, a method is required that is able to classify the risk level of dental diagnosis in order to assess the appropriate treatment. The study aims to apply the *Naïve Bayes* algorithm with the *Knowledge Discovery in Database (KDD)* methodology to classify dental diagnoses based on severe, moderate and mild risks. Testing of *naïve bayes* algorithm implementation was conducted with nine *k-fold cross-validation* scenarios. The use of *3-fold cross-validation* showed the best performance with 93.33% accuracy and kappa value of 0.855 which belongs to the *Very Good* classification category. While the average test result obtained accuracy 92.32% and kappa value of 0.833 which belongs to the category of classification *Very Good*. Thus, *Naïve Bayes* algorithm is considered able to classify the level of risk of dental diagnosis that occurs in UPTD Cingambul Public Health Center well because it has a good accuracy value and kappa value that belongs to the *Very Good* category.

Keywords: *cross-validation, dental diagnosis, KDD, naïve bayes*

1. PENDAHULUAN

Penyakit gigi merupakan kondisi yang dialami ketika muncul rasa sakit pada organ sekitar gigi dan mulut yang disebabkan karena gigi berlubang atau hal lainnya. Kesehatan gigi sangatlah penting karena gigi merupakan salah satu organ sistem pencernaan tempat menghancurkan makanan, jika terjadi gangguan dapat menyebabkan masalah kesehatan lain. Bahkan pada kondisi yang parah jika tidak segera ditangani peradangan pada gigi dan gusi dapat menyebar sampai jaringan periodontal yang menyebabkan penyakit mematikan seperti stroke, jantung dan diabetes [1]. Penyakit gigi menimbulkan ketidaknyamanan yang seringkali dikeluhkan orang-orang. Berdasarkan data *The Global Burden of Disease Study* pada tahun 2017 hampir 3,5 miliar penduduk di dunia mengalami penyakit gigi dan mulut. Diperkirakan secara global sebanyak 2,3 miliar penduduk mengalami karies gigi permanen dan lebih dari 530 juta anak menderita karies gigi sulung [2]. Di Indonesia berdasarkan hasil Riset Kesehatan Dasar (RIKESDAS) yang dilakukan Kementerian Kesehatan pada tahun 2018 sebanyak 57,6% penduduk Indonesia mengalami penyakit gigi, namun baru 10,2% yang mendapat penanganan dari tenaga medis [3]. Hal ini menunjukkan bahwa pelayanan kesehatan gigi di Indonesia masih rendah.

Begitu pula dengan pelayanan kesehatan gigi di UPTD Puskesmas Cingambul yang masih rendah. Karena rasio antara tumpatan dan pencabutan gigi pada tahun 2016 hanya mencapai rasio 0,02, pada tahun 2017 dengan rasio 0,45 dan pada tahun 2018 dengan rasio 0,32. Berdasarkan rekomendasi Kementerian Kesehatan indikator pelayanan kesehatan gigi yang baik ketika rasio antara pencabutan dan tumpatan gigi mencapai 1:1 [4]. Pada tahun 2020 di Poli Gigi UPTD Puskesmas Cingambul tercatat sebanyak 496 kunjungan. Dengan diagnosis karies gigi sebanyak 15 kasus, penyakit pulpa dan jaringan periapikal sebanyak 337 kasus, gingivitis dan jaringan periodontal sebanyak 47 kasus, kelainan dentofasial dan maloklusi sebanyak 3 kasus, serta gangguan gigi dan jaringan sebanyak 94 kasus. Diagnosis penyakit pulpa dan jaringan periapikal menjadi salah satu diagnosis dengan kondisi parah yang sering dikeluhkan di Poli Gigi UPTD Puskesmas Cingambul. Namun, yang menjadi masalah di UPTD Puskesmas Cingambul pelayanannya masih sangat terbatas karena tidak ada dokter yang bertugas. Sehingga hanya menangani pelayanan dasar saja seperti pencabutan ringan, penambalan atau pemberian antibiotik. Ketika ada suatu diagnosis dengan kondisi yang parah akan dirujuk ke rumah sakit untuk mendapatkan penanganan yang tepat. Oleh karena itu, dibutuhkan sebuah metode yang mampu mengklasifikasikan tingkat risiko berbagai diagnosis gigi yang terjadi di UPTD Puskesmas Cingambul agar dapat segera ditangani dengan tindakan yang tepat. Salah satu

caranya dengan menganalisis data riwayat kunjungan Poli Gigi UPTD Puskesmas Cingambul menggunakan *data mining* metode klasifikasi.

Data mining adalah proses penambangan informasi terpendam dalam sebuah *database* yang sebelumnya tidak diketahui [5]. Metode klasifikasi merupakan metode yang mengekstraksi model dengan mengelompokkan data latih pada sebuah kelas [6]. Metode ini telah dipakai pada beberapa penelitian sebelumnya untuk mendiagnosis berbagai jenis penyakit. Penelitian dilakukan untuk mendiagnosis penyakit *tuberculosis* dengan mengkomparasikan beberapa algoritma klasifikasi yaitu C4.5 yang memperoleh *accuracy* 84,56% dan nilai AUC 0,938, *Naïve Bayes* memperoleh *accuracy* 94,18% dan nilai AUC 0,977 serta *Neural Network* yang memperoleh *accuracy* 89,89% dan nilai AUC 0,975 [7]. Kemudian, algoritma *Naïve Bayes* dipakai untuk mengklasifikasikan penyakit *diabetes mellitus* yang memperoleh *accuracy* 90,20% dan nilai AUC 0,95 yang termasuk *Excellent Classification* [8]. Selain itu, algoritma klasifikasi juga dipakai untuk memprediksi penyakit jantung diantaranya *Naïve Bayes* dengan *accuracy* 83% dan nilai AUC 0,9, *Neural Network* dengan *accuracy* 81% dan nilai AUC 0,89 sedangkan *Random Forest* memperoleh *accuracy* 82% dan nilai AUC 0,89 [9]. Sehingga penggunaan algoritma *Naïve Bayes* dinilai memiliki performa yang baik dalam mengklasifikasikan berbagai jenis penyakit karena memperoleh *accuracy* maupun AUC yang cukup tinggi. *Naïve Bayes* merupakan pengklasifikasian yang menggunakan teknik probabilitas untuk memprediksi kejadian di masa depan berdasarkan kemiripan di masa lalu, dengan kelebihan dapat menangani atribut numerik maupun kategorikal, memiliki kecepatan waktu yang tinggi dan tidak memerlukan data latih yang besar [10].

Berdasarkan latar belakang tersebut, penelitian ini dilakukan untuk mengklasifikasikan tingkat risiko diagnosis gigi yang terjadi di UPTD Puskesmas Cingambul berdasarkan pada risiko parah, sedang dan ringan. Dengan tujuan untuk meningkatkan pelayanan kesehatan gigi di UPTD Puskesmas Cingambul dengan cara menentukan tingkat risiko dari setiap diagnosis yang terjadi agar dapat menilai penanganan yang tepat.

2. METODE PENELITIAN

Penelitian dilakukan dengan menggunakan data riwayat kunjungan Poli Gigi UPTD Puskesmas Cingambul pada tahun 2020 sebanyak 496 *records* yang terdiri dari 12 atribut yaitu Nama, Alamat, Jenis Kelamin, Usia, Keluhan Kondisi Gigi, Sisa Akar, Gusi Bengkak, Gigi Tumbuh, Skala Nyeri, Diagnosis, Tindakan dan Observasi. Metodologi yang dipakai pada penelitian ini adalah *Knowledge Discovery in Database* (KDD) yang merupakan sebuah pendekatan analisis untuk memodelkan kumpulan data menjadi pengetahuan yang berguna dengan

melakukan tahap-tahap pengurangan dimensi sehingga pola yang dibangun dinilai lebih efektif [11]. KDD terdiri dari lima tahap yaitu :

1. *Data Selection*

Tidak semua atribut dari data yang telah terkumpul digunakan. Pada tahap ini dilakukan pemilihan atribut relevan untuk klasifikasi tingkat risiko diagnosis gigi yaitu Jenis Kelamin, Usia, Keluhan Kondisi Gigi, Sisa Akar, Gusi Bengkak, Gigi Tumbuh, Skala Nyeri, Diagnosis, Tindakan dan Observasi. Karena metode klasifikasi termasuk *supervised learning* yang membutuhkan label pada data latihnya [6], maka atribut Observasi dipilih sebagai label yang terdiri dari tiga *class* yaitu parah, sedang dan ringan.

2. *Data Preprocessing*

Proses yang dilakukan pada tahap ini diantaranya memeriksa *missing value*, *inkonsisten data* atau data yang tidak relevan. Dari data yang telah terkumpul masih terdapat beberapa *missing value*, sehingga dilakukan penanganan dengan menghapus *records* yang memiliki *missing value* karena tidak memberikan informasi jika tetap dipertahankan [12]. Dari total 496 *records* sebanyak 46 *records missing value* dihapus dan tersisa 450 *records* saja.

3. *Data Transformation*

Tahap ini dilakukan untuk mengubah atau mentransformasikan data menjadi simbol atau *coding* agar sesuai dengan algoritma yang dipakai untuk mempermudah proses penambangan informasi. Beberapa algoritma dirancang untuk dapat menangani atribut numerik dan kategorikal, namun pada algoritma klasifikasi atribut kategorikal dianggap lebih bermanfaat pada tahap penambangan [13]. Oleh karena itu, atribut Usia yang memiliki tipe numerik diubah menjadi 7 interval kategori atau *bins* dengan menggunakan metode *Discretization by Binning* menggunakan *RapidMiner*. Sebelumnya ditentukan parameter batas minimal 1 dan maksimal 80 sesuai dengan data yang diperoleh. Hasil dari tahap transformasi data dengan *Discretization by Binning* terlihat pada Tabel 1 :

Tabel 1. Hasil Transformasi dengan *Discretization by Binning*

| Interval | Keterangan | Inisialisasi |
|----------|-------------|--------------------|
| Range 1 | 1 – 12,3 | Usia 1 - 12 Tahun |
| Range 2 | 12,3 – 23,6 | Usia 13 - 23 Tahun |
| Range 3 | 23,6 – 34,9 | Usia 24 - 34 Tahun |
| Range 4 | 34,9 – 46,1 | Usia 35 - 46 Tahun |
| Range 5 | 46,1 – 57,4 | Usia 47 - 57 Tahun |
| Range 6 | 57,4 – 68,7 | Usia 58 - 68 Tahun |
| Range 7 | 68,7 - 80 | Usia 69 - 80 Tahun |

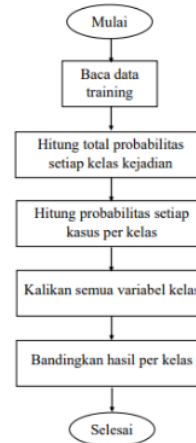
4. *Data Mining*

Pada tahap ini dilakukan penerapan algoritma *Naive Bayes* untuk mengklasifikasikan tingkat risiko diagnosis gigi pada UPTD Puskesmas

Cingambul. Algoritma *Naive Bayes* bekerja dengan melakukan pengklasifikasian berdasarkan nilai probabilitas pada setiap atributnya [10]. *Naive Bayes* memiliki persamaan ru [21] sebagai berikut :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Adapun tahap-tahap dari algoritma *Naive Bayes* seperti pada Gambar 1 di bawah ini :



Gambar 1. Tahap Algoritma *Naive Bayes*

Pengujian algoritma *Naive Bayes* pada penelitian ini dilakukan dengan menggunakan metode *k-fold cross-validation* dengan sembilan [3] skenario pengujian mulai dari *2-fold*, *3-fold*, *4-fold*, *5-fold*, *6-fold*, *7-fold*, *8-fold*, *9-fold* dan *10-fold*. *K-fold cross-validation* merupakan metode validasi yang membagi secara random seluruh data menjadi *data training* dan *data testing*, dengan keuntungan memanfaatkan data sebanyak mungkin untuk pelatihan dengan beberapa kali iterasi pengujian sesuai dengan nilai *k* yang ditentukan [14].

5. *Evaluation/interpretation*

Dari sembilan skenario pengujian yang telah dilakukan selanjutnya dievaluasi untuk menilai performa dari penggunaan algoritma *Naive Bayes* pada dataset Poli Gigi UPTD Puskesmas Cingambul berdasarkan nilai *accuracy* dari *confusion matrix* dan nilai *kappa*. Nilai *kappa* merupakan analisis statistik berdasarkan pada *class* yang berbeda, adapun klasifikasi nilai *kappa* menurut Altman (1991) yaitu sebagai berikut [15] :

- a. <0,2 : *Poor*
- b. 0,21-0,40 : *Fair*
- c. 0,41-0,60 : *Moderate*
- d. 0,61-0,80 : *Good*
- e. 0,81-1,00 : *Very Good*

3. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan untuk mengklasifikasikan tingkat risiko diagnosis gigi dengan menggunakan metodologi *Knowledge Discovery in Database* (KDD). Data yang digunakan adalah data riwayat kunjungan Poli Gigi UPTD Puskesmas Cingambul, setelah melewati tahap *data selection*, *data preprocessing* dan *data transformation* diperoleh sejumlah 450 records yang terdiri dari 9 atribut dengan 1 label/class. Beberapa tahap algoritma *Naïve Bayes* yaitu sebagai berikut :

1. Membaca *data training*

2. Menghitung probabilitas setiap kelas

Dari data yang digunakan terdapat 3 class dengan class parah berjumlah 319, sedang berjumlah 56 dan ringan berjumlah 75. Untuk menghitung probabilitas pada setiap kelasnya dengan menghitung jumlah setiap kelas dibagi dengan seluruh data. Probabilitas setiap kelas disajikan pada Tabel 2 di bawah ini :

Tabel 2. Probabilitas Setiap Kelas

| Kelas | Jumlah | Probabilitas Kelas P(H) | |
|--------|--------|-------------------------|--------|
| | | Parah | Sedang |
| Parah | 319 | 0,709 | |
| Sedang | 56 | 0,124 | |
| Ringan | 75 | 0,167 | |

3. Menghitung probabilitas setiap kejadian perkelas

Untuk menentukan nilai probabilitas setiap atribut dengan cara menghitung jumlah kejadian/atribut pada suatu kelas dibagi dengan kelas yang ada. Nilai probabilitas setiap atribut yaitu sebagai berikut :

a. Probabilitas atribut Jenis Kelamin seperti pada Tabel 3 :

Tabel 3. Probabilitas Atribut Jenis Kelamin

| Jenis Kelamin | Probabilitas P(XIH) | | |
|---------------|---------------------|--------|--------|
| | Parah | Sedang | Ringan |
| Perempuan | 0,702 | 0,714 | 0,573 |
| Laki-laki | 0,298 | 0,286 | 0,427 |

b. Probabilitas atribut Usia seperti pada Tabel 4 :

Tabel 4. Probabilitas Atribut Usia

| Usia | Probabilitas P(XIH) | | |
|--------------------|---------------------|--------|--------|
| | Parah | Sedang | Ringan |
| Usia 1 - 12 Tahun | 0,213 | 0,732 | 0,600 |
| Usia 13 - 23 Tahun | 0,197 | 0,071 | 0,067 |
| Usia 24 - 34 Tahun | 0,219 | 0,071 | 0,093 |
| Usia 35 - 46 Tahun | 0,163 | 0,018 | 0,133 |
| Usia 47 - 57 Tahun | 0,107 | 0,036 | 0,040 |
| Usia 58 - 68 Tahun | 0,069 | 0,071 | 0,027 |
| Usia 69 - 80 Tahun | 0,031 | 0 | 0,040 |

c. Probabilitas atribut Keluhan Kondisi Gigi seperti pada Tabel 5 :

Tabel 5. Probabilitas Atribut Keluhan Kondisi Gigi

| Keluhan Kondisi Gigi | Probabilitas P(XIH) | | |
|----------------------|---------------------|--------|--------|
| | Parah | Sedang | Ringan |
| Berlubang | 0,856 | 0,232 | 0,587 |
| Goyang | 0,088 | 0,357 | 0,413 |
| Patah | 0,016 | 0,179 | 0 |
| Tidak Ada | 0,041 | 0,232 | 0 |

d. Probabilitas atribut Sisa Akar seperti pada Tabel 6 :

Tabel 6. Probabilitas Atribut Sisa Akar

| Sisa Akar | Probabilitas P(XIH) | | |
|-----------|---------------------|--------|--------|
| | Parah | Sedang | Ringan |
| Ya | 0,370 | 0,482 | 0,307 |
| Tidak | 0,630 | 0,518 | 0,693 |

e. Probabilitas atribut Gusi Bengkak seperti pada Tabel 7 :

Tabel 7. Probabilitas Gusi Bengkak

| Gusi Bengkak | Probabilitas P(XIH) | | |
|--------------|---------------------|--------|--------|
| | Parah | Sedang | Ringan |
| Ya | 0,282 | 0,143 | 0,133 |
| Tidak | 0,718 | 0,857 | 0,867 |

f. Probabilitas atribut Gigi Tumbuh seperti pada Tabel 8 :

Tabel 8. Probabilitas Gigi Tumbuh

| Gigi Tumbuh | Probabilitas P(XIH) | | |
|-------------|---------------------|--------|--------|
| | Parah | Sedang | Ringan |
| Ya | 0,022 | 0,661 | 0,133 |
| Tidak | 0,978 | 0,339 | 0,867 |

g. Probabilitas atribut Skala Nyeri seperti pada Tabel 9 :

Tabel 8. Probabilitas Atribut Skala Nyeri

| Skala Nyeri | Probabilitas P(XIH) | | |
|--------------|---------------------|--------|--------|
| | Parah | Sedang | Ringan |
| Sering Sakit | 0,125 | 0,036 | 0 |
| Sakit | 0,828 | 0,339 | 0,640 |
| Tidak Sakit | 0,047 | 0,625 | 0,360 |

h. Probabilitas atribut Diagnosis seperti pada Tabel 10 :

Tabel 10. Probabilitas Atribut Diagnosis

| Diagnosis | Probabilitas P(XIH) | | |
|--|---------------------|--------|--------|
| | Parah | Sedang | Ringan |
| Penyakit Pulpa dan Jaringan Periapikal | 0,812 | 0,179 | 0,480 |
| Gangguan Gigi dan Jaringan | 0,060 | 0,750 | 0,360 |
| Gingivitis dan Jaringan Periodontal | 0,125 | 0,071 | 0,027 |
| Karies Gigi | 0 | 0 | 0,120 |
| Kelainan Dentofasial dan Maloklusi | 0,003 | 0 | 0,013 |

- i. Probabilitas atribut Tindakan seperti pada Tabel 11 :

Tabel 11. Probabilitas Atribut Tindakan

| Tindakan | Probabilitas P(X _{IIH}) | | |
|-------------|-----------------------------------|--------|--------|
| | Parah | Sedang | Ringan |
| Permedikasi | 0,925 | 0 | 0,013 |
| Pencabutan | 0 | 1 | 0,787 |
| Rujuk | 0,075 | 0 | 0 |
| Penambalan | 0 | 0 | 0,200 |

4. Mengalikan semua variabel perkelas
 Dari nilai probabilitas yang telah diperoleh bisa digunakan untuk memprediksi kejadian pada sebuah *data testing* seperti pada Gambar 2 dengan cara mengalikan setiap atribut perkelas.

| Jenis Kelamin | Usia | Keluhan Kondisi Gigi | Sisa Akar | Gusi Bengkak | Gigi Tumbuh | Skala Nyeri | Diagnosis | Tindakan | Observasi |
|---------------|--------------------|----------------------|-----------|--------------|-------------|-------------|--|-------------|-----------|
| Perempuan | Usia 69 – 80 Tahun | Berlubang | Ya | Ya | Tidak | Sakit | Penyakit Pulpa dan Jaringan Periapikal | Permedikasi | ? |

Gambar 2. Contoh Data Testing

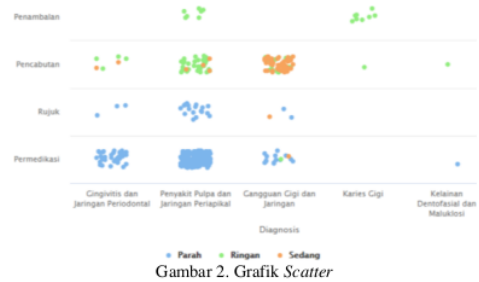
- a. Kelas Parah
 $= P(\text{Parah}) \times P(\text{Jenis Kelamin} = \text{Perempuan} | \text{Parah}) \times P(\text{Usia} = \text{Usia 69 – 80 Tahun} | \text{Parah}) \times P(\text{Keluhan Kondisi Gigi} = \text{Berlubang} | \text{Parah}) \times P(\text{Sisa Akar} = \text{Ya} | \text{Parah}) \times P(\text{Gusi Bengkak} = \text{Ya} | \text{Parah}) \times P(\text{Gigi Tumbuh} = \text{Tidak} | \text{Parah}) \times P(\text{Skala Nyeri} = \text{Sakit} | \text{Parah}) \times P(\text{Diagnosis} = \text{Penyakit Pulpa dan Jaringan Periapikal} | \text{Parah}) \times P(\text{Tindakan} = \text{Permedikasi} | \text{Parah})$
 $= 0,709 \times 0,702 \times 0,031 \times 0,856 \times 0,370 \times 0,282 \times 0,978 \times 0,828 \times 0,812 \times 0,925$
 $= 0,00084$
- b. Kelas Sedang
 $= P(\text{Sedang}) \times P(\text{Jenis Kelamin} = \text{Perempuan} | \text{Sedang}) \times P(\text{Usia} = \text{Usia 69 – 80 Tahun} | \text{Sedang}) \times P(\text{Keluhan Kondisi Gigi} = \text{Berlubang} | \text{Sedang}) \times P(\text{Sisa Akar} = \text{Ya} | \text{Sedang}) \times P(\text{Gusi Bengkak} = \text{Ya} | \text{Sedang}) \times P(\text{Gigi Tumbuh} = \text{Tidak} | \text{Sedang}) \times P(\text{Skala Nyeri} = \text{Sakit} | \text{Sedang}) \times P(\text{Diagnosis} = \text{Penyakit Pulpa dan Jaringan Periapikal} | \text{Sedang}) \times P(\text{Tindakan} = \text{Permedikasi} | \text{Sedang})$
 $= 0,124 \times 0,714 \times 0 \times 0,232 \times 0,482 \times 0,143 \times 0,339 \times 0,339 \times 0,179 \times 0$
 $= 0$
- c. Kelas Ringan
 $= P(\text{Ringan}) \times P(\text{Jenis Kelamin} = \text{Perempuan} | \text{Ringan}) \times P(\text{Usia} = \text{Usia 69 – 80 Tahun} | \text{Ringan}) \times P(\text{Keluhan Kondisi Gigi} = \text{Berlubang} | \text{Ringan}) \times P(\text{Sisa Akar} = \text{Ya} | \text{Ringan}) \times P(\text{Gusi Bengkak} = \text{Ya} | \text{Ringan}) \times P(\text{Gigi Tumbuh} = \text{Tidak} | \text{Ringan}) \times P(\text{Skala Nyeri} = \text{Sakit} | \text{Ringan}) \times P(\text{Diagnosis} = \text{Penyakit Pulpa dan Jaringan Periapikal} | \text{Ringan}) \times P(\text{Tindakan} = \text{Permedikasi} | \text{Ringan})$

$$= 0,167 \times 0,573 \times 0,040 \times 0,587 \times 0,307 \times 0,133 \times 0,867 \times 0,640 \times 0,480 \times 0,013$$

$$= 0,0000003$$

5. Membandingkan hasil setiap kelas
 Selanjutnya, hasil perhitungan manual setiap kelas dibandingkan yang mana nilai probabilitas terbesar adalah hasil klasifikasinya. Kelas parah memiliki nilai probabilitas terbesar 0,00084 sehingga *data testing* diatas memiliki tingkat risiko parah.

Hasil klasifikasi tingkat risiko diagnosis gigi dengan menggunakan algoritma *Naïve Bayes* disajikan pada Gambar 3 berupa grafik *scatter* di bawah ini :



Gambar 2. Grafik Scatter

Berdasarkan grafik *scatter* pada Gambar 2 penyakit pulpa dan jaringan periapikal menjadi diagnosis yang diklasifikasikan memiliki tingkat risiko parah paling tinggi yang sering terjadi dengan tindakan penanganan mulai dari permedikasi dan rujuk. Gangguan gigi dan jaringan diklasifikasikan sebagai diagnosis yang memiliki tingkat risiko sedang dengan tindakan penanganan pencabutan, rujuk, dan permedikasi. Sedangkan karies gigi pada setiap kasusnya diklasifikasikan memiliki tingkat risiko ringan yang masih dapat ditangani dengan tindakan penambalan.

Pengujian algoritma *Naïve Bayes* pada penelitian ini dilakukan dengan menggunakan sembilan skenario *k-fold cross-validation* untuk mengetahui performa terbaik dari model yang dihasilkan. Hasil evaluasi dari sembilan pengujian yang dilakukan disajikan pada Tabel 12 berikut :

Tabel 12. Perbandingan Hasil Evaluasi

| Skenario | Accuracy | Kappa |
|------------------|---------------|--------------|
| 2-fold | 91,56% | 0,816 |
| 3-fold | 93,33% | 0,855 |
| 4-fold | 91,78% | 0,821 |
| 5-fold | 92,67% | 0,84 |
| 6-fold | 92% | 0,826 |
| 7-fold | 92,89% | 0,844 |
| 8-fold | 91,78% | 0,821 |
| 9-fold | 92,44% | 0,835 |
| 10-fold | 92,44% | 0,835 |
| Rata-rata | 92,32% | 0,833 |

Berdasarkan hasil evaluasi pada Tabel 12, skenario dengan *3-fold cross-validation* menunjukkan performa terbaik dengan nilai *accuracy* 93,33% dan nilai *kappa* 0,855 yang termasuk kategori *Very Good*.

Sedangkan skenario dengan 2-fold cross-validation memperoleh nilai *accuracy* dan nilai *kappa* yang rendah yaitu 91,56% dan 0,816 yang masih termasuk kategori *Very Good*. Lalu, rata-rata hasil evaluasi dari sembilan skenario yang dilakukan diperoleh nilai *accuracy* 92,32% dan nilai *kappa* 0,833 yang termasuk kategori *Very Good*. Sehingga, algoritma *Naïve Bayes* dinilai mampu melakukan klasifikasi dengan baik pada dataset Poli Gigi UPTD Puskesmas Cingambul karena memiliki nilai *accuracy* dan nilai *kappa* yang cukup tinggi.

13

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, diperoleh beberapa kesimpulan yaitu :

1. Metodologi KDD dengan algoritma *Naïve Bayes* digunakan untuk mengklasifikasikan tingkat risiko diagnosis gigi yang terjadi di UPTD Puskesmas Cingambul berdasarkan risiko parah, sedang dan ringan. Algoritma *Naïve Bayes* bekerja dengan melakukan pembelajaran menggunakan nilai probabilitas pada setiap atributnya. Penyakit pulpa dan jaringan periapikal diklasifikasikan memiliki tingkat risiko parah paling tinggi dengan tindakan permedikasi dan rujuk. Sedangkan gangguan gigi dan jaringan diklasifikasikan memiliki tingkat risiko sedang, serta karies gigi diklasifikasikan memiliki tingkat risiko ringan.
2. Evaluasi penggunaan algoritma *Naïve Bayes* menunjukkan performa terbaik pada 3-fold cross-validation dengan memperoleh nilai *accuracy* 93,33% dan nilai AUC 0,855 yang termasuk kategori *Very Good*. Kemudian rata-rata dari sembilan skenario pengujian diperoleh nilai *accuracy* 92,32% dan nilai AUC 0,833 yang termasuk kategori *Very Good*. Sehingga algoritma *Naïve Bayes* dinilai mampu melakukan klasifikasi tingkat risiko diagnosis gigi dengan baik.

5. DAFTAR PUSTAKA

- [1] T. H. Saragih, W. F. Mahmudy, and Y. P. Anggodo. 2017. "Genetic Algorithm For Optimizing FIS Tsukamoto For Dental Disease Identification", *Int. Conf. Adv. Comput. Sci. Inf. Syst.*, pp. 345–349.
- [2] World Health Organization. 2020. "Oral Health" 2020. Available At: [https://www.who.int/news-room/factsheets/detail/oral-health] diakses 22 Februari 2021.
- [3] Kementerian Kesehatan. 2020. "Situasi Kesehatan Gigi dan Mulut 2019". Available At: https://www.kemkes.go.id/article/view/20030900005/situasi-kesehatan-gigi-dan-mulut-109] diakses 22 Februari 2021.
- [4] Dinas Kesehatan Provinsi Jawa Barat. 2017. Profil Kesehatan Dinas Kesehatan Provinsi Jawa Barat 2017: Dinas Kesehatan Provinsi Jawa Barat. Bandung.
- [5] I. Werdiningsih, B. Nuqoba, and Muhammadun. 2020. Data Mining Menggunakan Android, Weka, Dan SPSS: Airlangga University Press. Surabaya.
- [6] E. Prasetyowati. 2017. Data Mining Pengelompokan Data untuk Informasi dan Evaluasi: Duta Media Publishing . Pamekasan.
- [7] A. Amrin and H. Saiyar. 2019. "Aplikasi Diagnosa Penyakit Tuberculosis Menggunakan Algoritma Naive Bayes", *Jurikom*, vol. 5, no. 5, pp. 498–502.
- [8] A. Ridwan. 2020. "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus", *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21. Doi: 10.47970/siskom-kb.v4i1.169.
- [9] D. Derisma. 2020. "Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining", *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 84–88. Doi: 10.30871/jaic.v4i1.2152.
- [10] D. A. Pratiwi, R. M. Awangga, and M. Y. H. Setyawan. 2020. Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Teknik Informatika Menggunakan Metode Naïve Bayes: Kreatif. Bandung.
- [11] M. Arhami and M. Nasir. 2020. Data Mining Algoritma dan Implementasi: Penerbit Andi. Yogyakarta.
- [12] J. Gifari. 2020. "Teknik Pengolahan Data : Menenal Missing Values dan Cara-Cara Menanganinya". Available At : [https://www.dqlab.id/digital-transformation-pahami-teknik-pengolahan-ini-dalam-industri-data] diakses 3 Maret 2021.
- [13] P.-N. Tan, M. Steinbach, V. Kumar, and A. Karpatne. 2019. Introduction to Data Mining: Pearson Education. New York.
- [14] E. Anindika Sari, M. Thereza Br. Saragih, I. Ali Shariati, S. Sofya, R. Al Baihaqi, and R. Nooraeni. 2020. "Klasifikasi Kabupaten Tertinggal di Kawasan Timur Indonesia dengan Support Vector Mach", *JIKO (Jurnal Inform. dan Komputer)*, vol. 3, no. 3, pp. 188–195. Doi: 10.33387/jiko.v3i3.2364.
- [15] U. Enri. 2018. "Penerapan Algoritma C4.5 Dalam Pemilihan Program Studi Fakultas Ilmu Komputer (Studi Kasus : Sekolah Menengah Atas Negeri 1 Tambun Utara)", *J. Rekayasa Inf.*, vol. 7, no. 1, pp. 1-7.

ORIGINALITY REPORT

12%

SIMILARITY INDEX

10%

INTERNET SOURCES

5%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

| | | |
|---|---|-----|
| 1 | ejournal.unkhair.ac.id Internet Source | 2% |
| 2 | Submitted to Universitas Brawijaya Student Paper | 1% |
| 3 | www.sumobrain.com Internet Source | 1% |
| 4 | Submitted to Universitas Negeri Semarang Student Paper | 1% |
| 5 | jurnal.polibatam.ac.id Internet Source | 1% |
| 6 | ejournal.istn.ac.id Internet Source | 1% |
| 7 | dqlab.id Internet Source | 1% |
| 8 | journal.poltekkes-mks.ac.id Internet Source | 1% |
| 9 | Submitted to Anglia Ruskin University Student Paper | <1% |

| | | |
|----|---|------|
| 10 | journal.umpalangkaraya.ac.id Internet Source | <1 % |
| 11 | wrap.warwick.ac.uk Internet Source | <1 % |
| 12 | Maryam Hasan. "PREDIKSI TINGKAT KELANCARAN PEMBAYARAN KREDIT BANK MENGGUNAKAN ALGORITMA NAÏVE BAYES BERBASIS FORWARD SELECTION", ILKOM Jurnal Ilmiah, 2017 Publication | <1 % |
| 13 | doku.pub Internet Source | <1 % |
| 14 | eprints.binus.ac.id Internet Source | <1 % |
| 15 | 123dok.com Internet Source | <1 % |
| 16 | sinta.ristekbrin.go.id Internet Source | <1 % |
| 17 | Ilmianti Ilmianti, Indrya Kirana Mattulada, Sari Aldilawati, Sarahfin Aslan, Mila Febriany, Muh. Mursyid Hamka. "Media Komunikasi, Informasi dan Edukasi Terhadap Pengetahuan Anak Sekolah Dasar Tentang Kesehatan Gigi Mulut", Sinnun Maxillofacial Journal, 2021 Publication | <1 % |

repository.unpas.ac.id

18

Internet Source

<1 %

19

www.researchgate.net

Internet Source

<1 %

20

www.scribd.com

Internet Source

<1 %

21

Desi Ratna Sari, Dedy Hartama, Irfan Sudahri Damanik, Anjar Wanto. "Penerapan Metode Naive Bayes dalam Memprediksi Kepuasan Mahasiswa Terhadap Cara Pengajaran Dosen", Prosiding Seminar Nasional Riset Information Science (SENARIS), 2019

Publication

<1 %

22

Evita Fitri. "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine", Jurnal Transformatika, 2020

Publication

<1 %

23

Nuril Hidayanti, Dyah Titisari. "Low Cost Monitoring Kesehatan Berbasis Internet of Thing", Jurnal Teknokes, 2020

Publication

<1 %

24

authorzilla.com

Internet Source

<1 %

25

docplayer.net

Internet Source

<1 %

26

sites.google.com

Internet Source

<1 %

27

Eric Kasten. "", IEEE Transactions on Knowledge and Data Engineering, 4/2007

Publication

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off