

Penerapan Algoritma Naïve Bayes Untuk Memprediksi Diagnosa Penyakit Tuberculosis(TBC) Pada Puskesmas Umbansari

Frety Dwinda Aritonang¹, Loneli Costaner²

^{1,2}Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Lancang Kuning

^{1,2,3}Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015

e-mail: ¹fretydwindaart@gmail.com, ²lonelicostaner@unilak.ac.id

Abstrak

Algoritma naïve byaes merupakan salah satu metode data mining yang bertujuan untuk menemukan pola- pola yang baru dalam data dan menyaring dan menyaring data-data yang berasal dari kata yang telah ada pada masa sebelumnya. Puskesmas umbansari merupakan salah satu pusat layanan kesehatan masyarakat yang menangani beragam jenis penyakit setiap harinya salah satunya penyakit tuberculosis(tbc)metode yang digunakan adalah algoritma naïve bayes untuk memprediksi diagnosa penyakit tbc menggunakan 250 data set pasien dengan 9 gejala yang dialami oleh pasien. metode ini mengaplikasikan pengklasifikasi probabilistik sederhana, memanfaatkan asumsi bahwa semua atribut independen dan tidak saling terkait terhadap nilai kelas yang diberikan dengan akurasi mencapai 93.10% pada data uji. Hasil penelitian ini menunjukkan bahwa algoritma naïve bayes dapat digunakan secara efektif dala mendukung proses diagnosa tbc pada puskesmas umbansari ,serta memberikan prediksi yang akurat.

Kata Kunci: Algoritma naïve bayes, Puskesmas, Data Mining.

Abstract

The naïve byaes algorithm is a data mining method that aims to find new patterns in data and filter and filter data that comes from words that have existed in the past. Umbansari Community Health Center is one of the community health service centers that handles a variety of types of disease every day, one of which is tuberculosis (TB). The method used is the Naïve Bayes algorithm to predict the diagnosis of TB disease using 250 patient data sets with 9 symptoms experienced by the patient. This method applies a simple probabilistic classifier, utilizing the assumption that all attributes are independent and not related to the class value given with an accuracy of 93.10% on test data. The results of this research show that the Naïve Bayes algorithm can be used effectively to support the TB diagnosis process at the Umbansari Community Health Center, as well as providing accurate predictions.

Keywords: *Naive Bayes algorithm, community health center, Data Mining.*

1. PENDAHULUAN

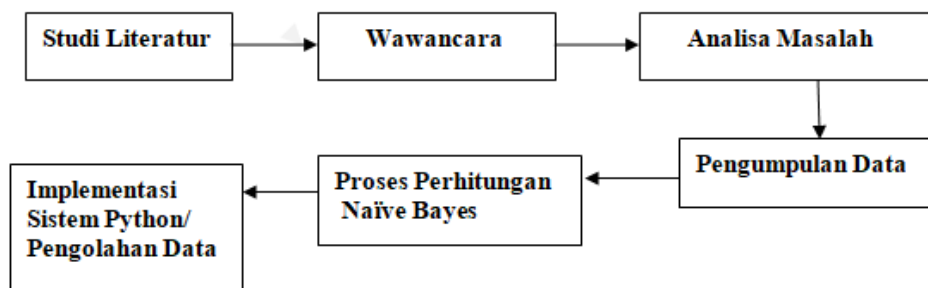
Perkembangan dan kemajuan teknologi di era globalisasi memberikan pengaruh yang cukup besar terhadap kehidupan manusia, baik berupa pendidikan, otomotif, dan lain-lain, terkhususnya pada dunia kesehatan. TBC atau Tuberculosis adalah salah satu penyakit menular yang dapat menginfeksi semua kalangan mulai dari bayi, anak-anak, remaja sampai lansia dan menimbulkan kesakitan dan kematian, penyakit Tuberculosis ini disebabkan oleh bakteri Mycobacterium Tuberculosis. Penyakit ini biasanya menginfeksi paru-paru, tetapi penyakit ini juga dapat menyerang organ lain yang ada di dalam tubuh manusia . Dalam data mining banyak algoritma yang dapat menyelesaikan sebuah permasalahan maka dari itu peneliti menggunakan algoritma Naive Bayes untuk membantu menyelesaikan permasalahan yang terjadi di puskesmas umbansari dalam mendiagnosa penyakit TBC. Algoritma Naive Bayes merupakan pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset. Teorema bayes

mengasumsikan semua atribut independen serta tidak saling keterantungan terhadap nilai yang diberikan kepada kelas variabel. Metode ini hanya membutuhkan jumlah data training yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Saat ini Puskesmas Umbansari masih menggunakan alat seadanya dalam mendiagnosa penyakit TBC, yaitu melalui pemeriksaan fisik, analisis gejala, dan hasil tes laboratorium seperti tes dahak dan rontgen. Sistem ini sangat bergantung pada keahlian dan pengalaman tenaga medis, serta memerlukan waktu yang cukup lama untuk mendapatkan hasil yang akurat. Proses ini juga dapat menimbulkan beban kerja yang tinggi bagi tenaga medis, terutama ketika jumlah pasien yang harus ditangani banyak.

Pada saat wawancara peneliti melakukan pertanyaan kepada pihak Puskesmas, bagaimana teknik dalam melakukan pemeriksaan terhadap pasien dengan tahap diagnosa awal dalam mengetahui positif atau negatif penyakit TBC, untuk mendiagnosa penyakit TBC memerlukan waktu yang cukup lama. Perlu satu 1-4 minggu untuk mendapatkan proses hasil pemeriksaan laboratorium pada penyakit TBC serta menimbulkan beban kerja yang tinggi bagi tenaga medis, terutama ketika jumlah pasien yang harus ditangani banyak setelah mengikuti kemajuan teknologi sangat diharapkan pihak Puskesmas dapat menjadikan hasil diagnosa sebagai acuan untuk memprediksi diagnosa penyakit TBC Pada pasien di masa yang akan datang untuk mengatasi masalah keterlambatan diagnosa penyakit TBC di Puskesmas Umbansari, penelitian ini menawarkan solusi dengan menerapkan teknik data mining menggunakan algoritma Naïve Bayes untuk membantu pihak puskesmas untuk memprediksikan diagnosa penyakit TBC .

Dengan adanya permasalahan tersebut di dapatkanlah solusi untuk mengatasi masalah terhadap pasien untuk memprediksi diagnosa penyakit TBC maka penulis mengambil judul skripsi “Penerapan Algoritma Naïve bayes Untuk Memprediksi Diagnosa Penyakit Tuberculosis (TBC) Pada Puskesmas Umbansari”

2. METODE PENELITIAN



Gambar 1. Tahap-tahap Penelitian

Adapun penjelasan langkah langkah metode penelitian Memprediksi Diagnosa Penyakit Tuberculosis (TBC) dapat diuraikan sebagai berikut :

1. Studi literatur
Studi literatur adalah mengumpulkan bahan refensi terkait metode yang digunakan dalam penelitian yang berasal dari jurnal,buku, dan sumber lainnya.
2. Wawancara
Melakukan Tanya jawab langsung kepada ibu Juliance Amd.Kep , yaitu dengan mengadakan wawancara secara langsung di puskesmas Umban Sari Dengan menanyakan gejala apa saja yang dialami oleh pasien.
3. Analisa Masalah
Analisis masalah dilakukan untuk mengidentifikasi masalah dan kebutuhan yang akan digunakan dalam melakukan pemecahan masalah, sebelumnya terlebih dahulu dilakukan beberapa tahap analisis untuk mengidentifikasi segala

kebutuhan yang akan diterapkan dalam agar tidak terjadi kesalahan. Dengan demikian, diharapkan penulis dapat menemukan kendala-kendala dan permasalahan yang terjadi pada Puskesmas Umban Sari dalam melakukan pemeriksaan terhadap

pasien sehingga peneliti dapat mencari solusi dari permasalahan tersebut.

Sistem yang akan dibangun adalah bagaimana agar pasien dapat mengetahui gejala awal pada penyakit yang dideritanya sebelum langsung kerumah sakit.

4. Pengumpulan Data

Pada tahap ini dilakukan proses pengumpulan data dengan metode studi literatur, wawancara dan observasi untuk melakukan pengamatan dan analisa terhadap proses pemeriksaan kesehatan pada Puskesmas Umban Sari sehingga mendapatkan data dan informasi yang dibutuhkan oleh peneliti

5. Proses perhitungan algoritma naïve bayes.

Pada tahap ini dilakukanlah proses perhitungan manual dan dengan menggunakan Metode naïve bayes, untuk mencari nilai akurasi yang baik yang mencapai pendekatan 100%. Sistem yang berjalan

untuk pengambilan keputusan terhadap pasien yang terdiagnosa penyakit tuberculosis (TBC) saat ini masih dikerjakan secara manual.

6. Implementasi Sistem Denga python/Pengolahan Data

Implementasi sistem merupakan tahap untuk merealisasikan hasil yang telah dilakukan sebelumnya. Pada tahap ini, penulis melakukan proses pengujian dengan menggunakan software python untuk memprediksi yang berguna untuk mendiagnosa penyakit pada pasien yang akan melakukan pemeriksaan kesehatan pada puskesmas Umban Sari dengan menggunakan metode Naïve Bayes.

3. HASIL DAN PEMBAHASAN

Pada bagian ini, menguraikan hasil penelitian dan membahas temuan utama yang diperoleh. Analisis dilakukan terhadap data yang dikumpulkan untuk penelitian, serta mengevaluasi performa metode yang digunakan. Hasil dan interpretasi dari analisis ini akan diuraikan dalam beberapa sub-bagian.

3.1 Analisa Masalah

Analisis masalah dilakukan untuk mengidentifikasi secara spesifik kendala-kendala yang dihadapi dalam proses diagnosa TBC di Puskesmas Umbansari. Dari analisis ini, terungkap beberapa masalah seperti ketidakpastian dalam diagnosa berdasarkan gejala, kebutuhan akan pendekatan yang lebih sistematis dalam penggunaan data pasien

3.2 Menentukan Metode

Pemilihan metode Naïve Bayes sebagai algoritma utama didasarkan pada keunggulannya dalam mengatasi kompleksitas diagnosa berbasis data, khususnya dalam mengidentifikasi TBC dari sekumpulan gejala yang beragam.

3.3 Pengumpulan Data

Proses pengumpulan data dilaksanakan dengan mengakses catatan medis pasien dan basis data kesehatan Puskesmas Umbansari dengan jumlah data 250 data dari tahun 2022,2023,2024 dan 9 variabel yaitu keringat dingin,batuk,sesak nafas,sakit dada,lelah yang berlebihan demam,hilang nafsu makan,berat badan menurun,keringat di malam hari. Data yang terkumpul mencakup informasi penting seperti gejala yang dialami pasien, riwayat medis, hasil tes laboratorium

3.4 Proses Perhitungan Naïve Bayes

Melakukan perhitungan manual dengan metode naïve bayes dengan mengitung jumlah likelihood ya dan likelihood tidak peneliti mengambil 3 sampel degan rumus dibawah ini :

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

Dengan:

X = Data kelas yang belum diketahui;
H = Hipotesis data X merupakan suatu label kelas tertentu;
P(H|X) = Probabilistik hipotesis H berdasarkan kondisi X;
P(X|H) = Probabilistik X berdasarkan hipotesis H;
P(X) = Probabilistik X;

TABEL 1. Sample data

Nama	Keringat dingin	Batuk	Sesak nafas	Sakit dada	Lelah yang berlebihan	Demam	Hilang nafsu makan	Berat badan menurun	Keringat Di malam hari	Status
Ade widodi	-	-	Ada	Ada	-	Ada	-	Ada	Ada	Positif
juni	-	Ada	Ada	-	-	-	Ada	Ada	-	Negatif
indah	Ada	Ada	Ada	-	Ada	Ada	Ada	Ada	Ada	Positif

$$P(\text{TBC-Positif}) = \frac{\text{Jumlah positif}}{\text{total sampel data}} = \frac{2}{3} = 0,67$$

$$P(\text{TBC-Negatif}) = \frac{\text{Jumlah negatif}}{\text{Total sampel data}} = \frac{1}{3} = 0,33$$

3.5 Implementasi python

Dalam mengimplementasikan naïve bayes dalam pemrograman python dapat dilakukan dengan langkah langkah sebagai berikut;

1. Mengimport dataset

```
import pandas as pd
pd.set_option("display.max_rows", 10)

df = pd.read_excel('data_pasien_tbc.xlsx') # Membaca data dari file Excel
df # Menampilkan DataFrame awal
```

Gambar 3.1 Tahap membaca file excel

2. Menampilkan Datset TBC

Setelah memasukan dataset dan membuat data frame, peneliti melakukan penampilan dataset pasien TBC dan akan tampil data yang sudah di panggil tadi dengan source code dibawah ini :

0	1	2022	ADE WIDODI	Sri Meranti	38	L	Tidak	Ada	Tidak	Ada	Ada	Tidak	Ada	Tidak	Ada	Positif
1	2	2022	Afrizal	Umban Sari	50	L	Ada	Ada	Tidak	Ada	Tidak	Ada	Tidak	Ada	Ada	Positif
2	3	2022	akmal	Palas	38	L	Ada	Ada	Tidak	Ada	Tidak	Ada	Tidak	Ada	Ada	Positif
3	4	2022	ALFONCO HENDRIKUS HUTAURUK	Umban Sari	21	L	Ada	Tidak	Tidak	Ada	Ada	Ada	Tidak	Ada	Ada	Positif
4	5	2022	alpian caniago	Umban Sari	50	L	Tidak	Ada	Ada	Tidak	Tidak	Tidak	Tidak	Ada	Ada	Positif
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
245	246	2024	SUMIATI	Palas	47	P	Tidak	Ada	Tidak	Ada	Ada	Ada	Tidak	Tidak	Ada	Positif
246	247	2024	SYARIL RAMADANI	Sri Meranti	34	L	Ada	Tidak	Tidak	Ada	Ada	Ada	Ada	Ada	Ada	Positif
247	248	2024	WAN ZEYANIA ALYSSA	Umban Sari	32	P	Tidak	Ada	Ada	Tidak	Ada	Tidak	Ada	Ada	Ada	Positif
248	249	2024	YONDEDI	-	16	L	Ada	Tidak	Ada	Tidak	Tidak	Tidak	Ada	Ada	Tidak	Positif
249	250	2024	YUDA BRAMANTA PI ITPA	Sri Meranti	33	L	Ada	Ada	Ada	Tidak	Ada	Ada	Ada	Ada	Ada	Positif

Gambar 3.2 Melihat isi Data

3. Pemilihan fitur

Kodingan ini menghapus kolom-kolom yang tidak diperlukan dari DataFrame df seperti no,tahun,nama lengkap pasien,kelurahan,umur Dengan langkah ini, data sudah siap untuk proses analisis lebih lanjut berikut adalah souce code dan hasil setelah menghapus fitur yang tidak diperlukan.

```
df1 = pd.DataFrame(df.drop(['No','Tahun','Nama Lengkap Pasien','Kelurahan','Umur'], axis=1)) # Menghapus kolom yang tidak diperlukan dan menyimpan h
df1 # Menampilkan DataFrame setelah penghapusan kolom
```

	Jenis Kelamin	Keringat Dingin	Batuk	Sesak Nafas	Sakit Dada	Lelah Yang berlebihan	Demam	Hilang Nafsu Makan	Berat Badan Menurun	Keringat Di Malam Hari	Status
0	L	Tidak	Ada	Tidak	Ada	Ada	Tidak	Ada	Tidak	Ada	Positif
1	L	Ada	Ada	Tidak	Ada	Tidak	Ada	Tidak	Ada	Ada	Positif
2	L	Ada	Ada	Tidak	Ada	Tidak	Ada	Ada	Tidak	Ada	Positif
3	L	Ada	Tidak	Tidak	Ada	Ada	Ada	Ada	Tidak	Ada	Positif
4	L	Tidak	Ada	Ada	Tidak	Tidak	Tidak	Tidak	Tidak	Ada	Positif
...	...	...	...	...	...	...	...	...	...	...	...
245	P	Tidak	Ada	Tidak	Ada	Ada	Ada	Tidak	Tidak	Ada	Positif
246	L	Ada	Tidak	Tidak	Ada	Ada	Ada	Ada	Ada	Ada	Positif
247	P	Tidak	Ada	Ada	Tidak	Ada	Tidak	Ada	Ada	Ada	Positif
248	L	Ada	Tidak	Ada	Tidak	Tidak	Tidak	Ada	Ada	Tidak	Positif
249	L	Ada	Ada	Ada	Tidak	Ada	Ada	Ada	Ada	Ada	Positif

250 rows x 11 columns

Gambar 3.3 Pemilihan fitur

4. Menampilkan Hasil Prediksi Naïve Bayes

Menampilkan hasil prediksi dari data sebelumnya untuk melatih apakah naïve bayes dapat memprediksi data tersebut dengan baik Membuat DataFrame df_pred dari data uji X_test. DataFrame ini digunakan untuk menampung hasil prediksi dengan nilai asli dari data yang telah di encode sebelumnya inverse_transform dari LabelEncoder LE_gejala untuk mengconversi kolom gejala yang telah di encode kembali ke nilai aslinya applylamda val LE_gejala inverse _transform(vall)berati setiap kolom dalam rentang yang ditentukan akan diterapkn fungsi inverse_transform inverse_transform dari label Encoder LE_JK untuk mengonversi kolom 'jenis kelamin'yang telah di encode kembali ke nilai aslinya inverse_transform dari LabelEncoder LE_status untuk mengonversi kolom status pada tahap data uji

y_test dari hasil prediksi y_pred_nb ke nilai aslinya mereset indeks pada dataframe df_pred untuk memastikan indeks berurutan dan dimulai dari 0 setelah

	Jenis Kelamin	Keringat Dingin	Batuk	Sesak Nafas	Sakit Dada	Lelah Yang berlebihan	Demam	Hilang Nafsu Makan	Berat Badan Menurun	Keringat Di Malam Hari	Status	Prediksi Status	Hasil Prediksi
0	L	Ada	Ada	Tidak	Tidak	Tidak	Ada	Ada	Ada	Tidak	Positif	Positif	Benar
1	P	Tidak	Ada	Ada	Ada	Ada	Ada	Tidak	Ada	Ada	Positif	Positif	Benar
2	P	Tidak	Ada	Ada	Tidak	Tidak	Tidak	Ada	Tidak	Ada	Negatif	Negatif	Benar
3	L	Tidak	Ada	Tidak	Ada	Tidak	Ada	Ada	Tidak	Ada	Positif	Positif	Benar
4	L	Ada	Ada	Ada	Ada	Ada	Tidak	Ada	Ada	Ada	Positif	Positif	Benar
...	...	...	...	...	...	...	...	...	...	...	...	...	...
i3	L	Ada	Ada	Ada	Tidak	Ada	Ada	Ada	Tidak	Ada	Positif	Positif	Benar
i4	P	Ada	Ada	Tidak	Tidak	Ada	Tidak	Ada	Tidak	Tidak	Positif	Positif	Benar
i5	L	Tidak	Ada	Ada	Tidak	Ada	Ada	Ada	Tidak	Tidak	Positif	Positif	Benar

```

# Mengonversi hasil encoding ke nilai asli
df_pred = pd.DataFrame(X_test)
df_pred[df_pred.columns[1:]] = df_pred[df_pred.columns[1:]].apply(lambda val: LE_gejala.inverse_transform(val))
df_pred['Jenis Kelamin'] = LE_JK.inverse_transform(df_pred['Jenis Kelamin'])
df_pred['Status'] = LE_status.inverse_transform(y_test)
df_pred['Prediksi Status'] = LE_status.inverse_transform(y_pred_nb)

# Reset nomor indeks pada DataFrame
df_pred = df_pred.reset_index(drop=True)

# Menambahkan kolom 'Hasil Prediksi' yang menunjukkan apakah prediksi benar atau salah
df_pred['Hasil Prediksi'] = df_pred.apply(lambda row: 'Benar' if row['Status'] == row['Prediksi Status'] else 'Salah', axis=1)

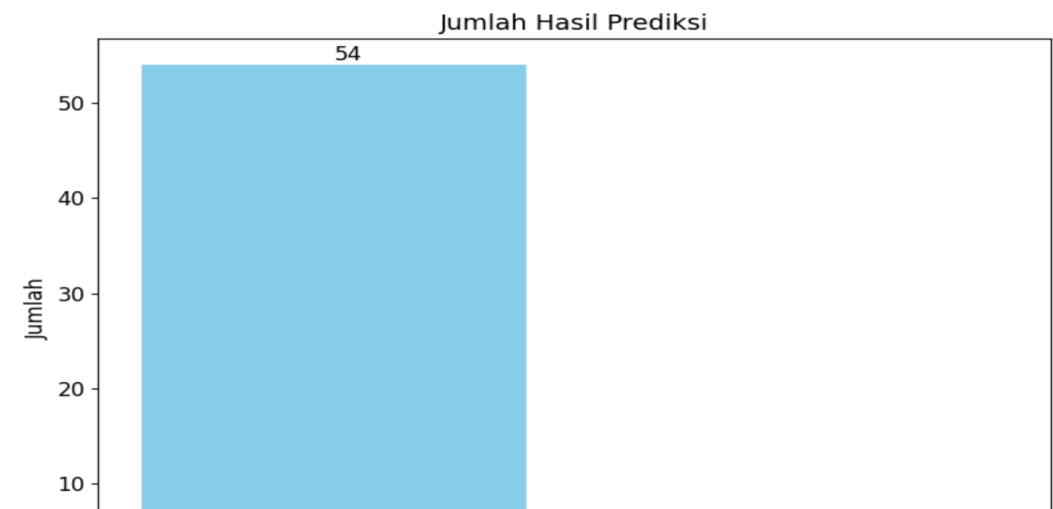
df_pred

```

Gambar 3.4 Hasil prediksi

5. Menampilkan diagram batang hasil prediksi

Setelah melakukan predisi dengan data test sebanyak 58 sehingga menghasilkan 54 hasil prediksi benar dan 4 hasil prediksi salah. Peneliti menampilkan diagram batang jumlah hasil prediksi dari kolom Hasil Prediksi dalam DataFrame df_pred. Ini adalah cara yang baik untuk memvisualisasikan distribusi hasil prediksi dalam bentuk grafis yang mudah dipahami



Gambar 3.5 Diagram hasil Prediksi

6. Menampilkan akurasi data training dan data testing

Pada penelitian ini peneliti membagi data menjadi dua yaitu data testing dan data training pembagian data nya adalah 70:30 menggunakan data training untuk melatih model dengan jumlah data 135 data dan data testing 58 data dijadikan untuk menguji model dalam memprediksi diagnosa penyakit

· Akurasi

```
] from sklearn.metrics import accuracy_score, classification_report
# Mengukur akurasi model dan menampilkan hasil
acc_nb_train = accuracy_score(y_train, y_pred_nb_train) * 100
acc_nb = accuracy_score(y_test, y_pred_nb) * 100 # Menghitung akurasi dikali 100

print(f'Akurasi Data Training\t: {acc_nb_train:.2f}%')
print(f'Akurasi Data Testing\t: {acc_nb:.2f}%')
```

→ Akurasi Data Training : 94.81%
Akurasi Data Testing : 93.10%

Gambar 3.6 Hasil Akurasi training dan testing

4. KESIMPULAN

Algoritma Naïve Bayes terbukti efektif untuk memprediksi diagnosis penyakit TBC pada Puskesmas Umbansari dengan akurasi 93.10%. Hasil ini menunjukkan bahwa algoritma tersebut dapat memanfaatkan data pasien untuk memberikan prediksi yang akurat dan dapat diandalkan dalam proses diagnosa penyakit TBC. Penerapan algoritma Naïve Bayes berhasil mengurangi waktu yang diperlukan untuk mendiagnosa penyakit TBC dari 1-4 minggu menjadi proses yang lebih cepat dan efisien. Dengan akurasi model yang tinggi, teknologi ini dapat membantu tenaga medis dalam mendiagnosa TBC secara lebih cepat dan mengurangi beban kerja mereka.

DAFTAR PUSTAKA

- [1] Aliful, A., Goejantoro, R., & Fathurahman, M. (n.d.). Klasifikasi Penyakit Tuberculosis Menggunakan Metode Naive Bayes (Studi Kasus: Data Pasien Di Puskesmas Petung Kabupaten Penajam Paser Utara) Classification of Tuberculosis Using the Naive Bayes Method (Case Study: Patient Data at Puskesmas Petung Penajam Paser Utara). *Jurnal EKSPONENSIAL*, 14(1). <http://jurnal.fmipa.unmul.ac.id/index.php/exponensial>
- [2] Bidang Komputer Sains dan Pendidikan Informatika, P., Akademi Perekam dan Informasi Kesehatan Iris Padang Jl Gajah Mada No, D., & Barat, S. (n.d.). *Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Yuli Mardi*.
- [3] Cherian, V., & Student, J. (n.d.). Heart Disease Prediction Using Naïve Bayes Algorithm and Laplace Smoothing Technique. *International Journal of Computer Science Trends and Technology (IJCST)*, 5. www.ijcstjournal.org
- [4] Manalu, E., Sianturi, F. A., & Manalu, M. R. (2015). (). *Jl. Iskandar Muda No.1 Medan*, 1(2), 1.
- [5] Marlina, E., Salman Halim, N., Donny, R., & Informasi STMIK Dipanegara Makassar, S. (n.d.). *Sistem Pakar Untuk Diagnosa Penyakit Vertigo Dengan Metode Forward Chaining (Studi Kasus: Klinik Bhayangkara Panaikang)*.
- [6] Mutiara, E. (2020). Algoritma Klasifikasi Naive Bayes Berbasis Parricel Swarm Optimization Untuk Prediksi Penyakit Tuberculosis (TB). *Jurnal Swabumi*, 8(1), 46–58.
- [7] Napirah, M. R., Rahman, A., & Tony, A. (2016). Faktor- Faktur Yang Berhubungan Dengan Pemanfaatan Pelayanan Kesehatan Di Wilayah Kerja Puskesmas Tambaran Kecamatan Poso Pesisir Utara Kabupaten Poso. *Jurnal Pengembangan Kota*, 4(1), 29. <https://doi.org/10.14710/jpk.4.1.29-39>
- [8] Ovilianda, A., & Ginting, B. (2021). Penerapan Data Mining Korelasi Penjualan Spare Part Mobil Menggunakan Metode Algoritma Apriori (Studi Kasus: CV. Citra Kencana Mobil). In *Journal Of Information And Technology Unimor*.

