

## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Ikhsan Nuh Attahalla, Adithia Jovandy, Hanif Ha Bibie (2018) dalam “Klasifikasi penyakit Kanker payudara Menggunakan metode *K Nearest Neighbor(KNN)* (Studi Kasus: Universitas Sriwijaya Palembang, Indonesia)”. Tujuan dari penelitian ini adalah dapat memprediksi jenis kanker, jinak atau ganas.

Zuriati Z1, Qomariyah N2(2023) dalam “Klasifikasi Penyakit Stroke Menggunakan Algoritma *K-Nearest Neighbor (KNN)* (Studi Kasus: Politeknik Negeri Lampung)”. Tujuan dari penelitian ini adalah untuk menerapkan algoritma KNN untuk klasifikasi data penyakit stroke.

Syahrani Lonang<sup>1</sup>, Dwi Normawati<sup>2</sup>(2021) dalam “Klasifikasi Status Pada Balita Menggunakan *K-nearest Neighbor(KNN)* Dengan Feature Selection *Backward Elimination* (Studi Kasus: Puskesmas Ubung, Kecamatan Jonggat, Kabupaten Lombok Tengah)”. Tujuan dari penelitian ini adalah membuat sebuah sistem yang dapat mengklasifikasi juga memeriksa status stunting pada balita dengan memanfaatkan algoritma *K-Nearest Neighbor* dengan seleksi fitur *Backward Elimination*.

Hasran,<sup>1</sup>(2020) dalam “Klarifikasi Penyakit Jantung Menggunakan Metode *K-Neighbor Neighbor(KNN)* (Studi Kasus: Lab. Riset Fakultas Ilmu Komputer Universitas Muslim Indonesia)”. Tujuan dari penelitian ini adalah untuk menghitung performa metode *K-Nearest Neighbor (KNN)* pada data penyakit *kardiovaskuler*.

Atika Putri Dewi<sup>1</sup>, Eryati Darwin<sup>2</sup>, Edison<sup>3</sup>(2013) dalam “Hubungan Tingkat Pengetahuan Ibu dengan Pemberian Imunisasi Dasar Lengkap pada Bayi di Kelurahan Parupuk Tabing Wilayah Kerja Puskesmas Lubuk Buaya Kota Padang Menggunakan Metode *Cross-sectional* ( Studi kasus: Kelurahan Parupuk Tabing Wilayah Kerja Puskesmas Lubuk Buaya Kota Padang)”. Tujuan dari penelitian ini adalah untuk mengetahui hubungan tingkat pengetahuan ibu dengan pemberian imunisasi dasar lengkap bayi di wilayah kerja Kelurahan Parupuk Tabing Wilayah Kerja Puskesmas Lubuk Buaya Kota Padang Tahun 2013.

Diusulkan : Aurea Imelda Mone(2024) Klasifikasi Efek Imunisasi Campak Pada bayi Dengan Metode *KNN* Berbasis Web ini , dapat membantu orang tua dalam menentukan kapan waktu yang tepat untuk melakukan imunisasi campak. Dan dengan dibuatnya aplikasi web kesehatan ini dapat membantu orang tua melakukan simulasi imunisasi campak pada bayi dan membantu memprediksi keluhan dan efek yang mungkin timbul setelah imunisasi .

Perbandingan dari penelitian ini dengan penelitian sebelumnya adalah adanya kesamaan dalam menggunakan *KNN*,Web, namun pada penelitian ini lebih difokuskan kepada prediksi efek imunisasi campak pada bayi.

*Tabel 2 1 Perbandingan Penelitian*

No	Penulis, Tahun	Objek	Metode	Teknologi	Hasil
1.	Ikhsan Nuh Atthalla, Adithia Jovandy,	Kanker Payudara Tingkat Ganas	<i>K-Nearet</i> <i>Neighbr</i>	web	Klasifikasi penyakit kanker payudara memperoleh hasil

	Hanifa ha bibie, 2018				paling akurat sebesar 93%
2.	Zuriati Z1, Qomariyah N2, 2023	Stroke	<i>K-Nearest Neighbor</i>	Mobile	Klasifikasi menggunakan algoritma KNN untuk dataset penyakit stroke telah berhasil dilakukan, nilai akurasi tertinggi didapatkan pada nilai $k=5$ dengan komposisi perbandingan data training dan testing 90% : 10% dengan nilai akurasi 93.54%.
3.	Syahrani Lonang1, Dwi Normawati2,2 021	Klasifikasi Stunting Pada Balita	<i>K-Nearest Neighbor</i>	Web	Tingkat akurasi yang didapatkan <i>k-nearest neighbor</i> dengan fitur seleksi <i>backward elimination</i> mencapai 92,2% meningkat 0,30%.
4.	Hasran,2020	Jantung	K-Nearest Neighbor	<i>Jupyter Notebook</i>	Dilakukan perhitungan peforma dari seluruh data testing dengan berbagai simulasi ketetangaan pada metode KNN

5.	Atika Putri Dewi1, Eryati Darwin2, Edison3, 2013	Pemberian Imunisasi Dasar Lengkap Pada Bayi	<i>Cross- sectiona al</i>	Web	bahwa lebih dari separuh responden memberikan imunisasi lengkap pada bayi dikelurahan Parupuk Tabing wilayah kerja Puskesmas Lubuk Buaya Kota Padang. Lebih dari separuh responden memiliki
			mengenai		pengetahuan yang cukup imunisasi dasar lengkap pada bayi dikelurahan Parupuk Tabing wilayah kerja Puskesmas Lubuk Buaya Kota Padang.
6.	Usulan	Klasifikasi Efek imunsasi campak pada bayi	<i>K-Nearest Neighbor</i>	-Web -Laravel -Python	Memprediksi atau mengklasifikasikan efek imunisasi pada balita, dan membuat suatu aplikasi simulasi imunisasi campak pada bayi dan memprediksi efek yang mungkin akan timbul jika melakukan imunasi campak dengan kondisi

					bayi yang sudah di tentukan.
--	--	--	--	--	------------------------------

## 2.2 Dasar Teori

### 2.2.1 Imunisasi

Imunisasi merupakan bentuk intervensi kesehatan yang sangat efektif dalam menurunkan angka kematian bayi. Dengan imunisasi, berbagai penyakit seperti *TBC, difteri, pertusis, tetanus, hepatitis B* dan lainnya dapat dicegah. Pentingnya imunisasi dapat dilihat dari banyaknya bayi yang meninggal akibat penyakit yang dapat dicegah dengan imunisasi (PD31). Tujuan imunisasi untuk memberikan perlindungan terhadap penyakit yang dapat dicegah dengan imunisasi. Menurut Permenkes RI (2017), program imunisasi di Indonesia memiliki tujuan umum untuk menurunkan angka kesakitan, kecacatan, dan kematian akibat Penyakit yang Dapat Dicegah Dengan Imunisasi (PD31).

Namun sampai saat ini masih banyak bayi yang belum di Imunisasi karena banyaknya efek samping yang timbul setelah melakukan Imunisasi seperti Rasa sakit atau kemerahan di tempat suntikan, Demam, kehilangan nafsu makan atau mengantuk, Reaksi elergi, Muntah-muntah dan masih banyak lagi. Timbulnya efek samping dari imunisasi tersebut dikarenakan kurangnya informasi tentang riwayat kesehatan bayi dan kurangnya pendampingan atau konsultasi terhadap Dokter atau Rekam medis.

Pemrosesan data dapat dilakukan dengan menggunakan *data meaning*. Data meaning adalah sebuah teknik pengolahan data dengan berbagai macam cara didalamnya. Pada *data mining*, proses dilakukan untuk menemukan informasi dari kumpulan data dan menemukan sebuah informasi baru yang tersimpan pada kumpulan data yang dapat dimanfaatkan pada proses pengambilan keputusan. Salah satu *algoritma* yang bisa digunakan untuk menyelesaikan permasalahan pada *data meaning* yaitu *algoritma K-Nearest Neighbor*. *KNN* sendiri merupakan bagian dari pada *algoritma* terpelajar (*supervised learning*) yang artinya pada dataset diharuskan memiliki target ataupun kelas.

Pada penelitian kali ini mengambil 1 jenis imunisasi yaitu Campak. Vaksin Campak adalah vaksin yang digunakan untuk mencegah penyakit campak. Vaksin campak terbuat dari virus campak yang dilemahkan, penyuntikan vaksin campak akan membuat tubuh memproduksi antibodi yang akan melawan virus tersebut jika sewaktu waktu menyerang. Adapun Efek samping dari jenis imunisasi ini yaitu Demam atau pusing, Kehilangan nafsu makan, Mual atau muntah, Nyeri otot,lelah dan Nyeri atau kemerahan diarea penyuntikan

### **2.2.2 Data Mining**

Secara sederhana data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004). *Data mining* juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2007). Data mining, sering juga disebut sebagai *knowledge discovery in database (KDD)*. *KDD* adalah

kegiatan yang meliputi pengumpulan, pemakaian data, *history* untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar(Santoso, 2007).

Karakteristik *data mining* sebagai berikut:

- a. *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- b. *Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih percaya.
- c. *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi (Davies, 2004).

### **2.2.3 Klasifikasi Data Mining**

Klasifikasi *data mining* adalah sebuah proses menemukan definisi kesamaan karakteristik dalam suatu kelompok atau kelas (*class*). Metode *klasifikasi* merupakan teknik yang didasarkan pada atribut dari kelompok yang sudah didefinisikan, sehingga didapatkan suatu aturan yang digunakan untuk melakukan *klasifikasi* pada data dengan cara memanipulasi data yang sudah ada dan sudah diklasifikasi. (Novriansyah & Nurcahyo, 2015).

Metode ini termasuk ke dalam kelompok *supervised learning* yang setiap item datanya memiliki label atau kelas yang pengaruhi *atribut*. Tipe data yang cocok digunakan pada metode *klasifikasi* yaitu *biner* atau *nominal* sedangkan untuk tipe data *ordinal* kurang cocok sebab pada metode ini menggunakan pendekatan secara *implisit*. (Novriansyah & Nurcahyo, 2015).

#### 2.2.4 K-Nearest Neighbor

Algoritma *K-Nearest Neighbor*(KNN) adalah algoritma *machine learning* yang bersifat *non-parametric* dan *lazy learning*. Metode yang bersifat *non-parametric* memiliki makna bahwa metode tersebut tidak membuat asumsi apa pun tentang distribusi data yang tetap dalam model, terlepas data tersebut berukuran kecil ataupun besar. Tujuan dari algoritma *k-nearest neighbor* adalah untuk mengidentifikasi tetangga terdekat dari titik *kueri* yang diberikan, sehingga kita dapat menetapkan label kelas ke titik tersebut.

##### 1. Menentukan *metrik* jarak

Untuk menentukan titik data mana yang paling dekat dengan titik *kueri* tertentu, jarak antara titik *kueri* dan titik data lainnya perlu dihitung. *Metrik* jarak ini membantu membentuk batasan keputusan, yang mengarahkan *kueri* partisi ke kelas yang berbeda. Untuk menemukan titik serupa terdekat, kita bisa menggunakan perhitungan jarak seperti *Euclidean distance*, *Hamming distance*, *Manhattan distance* dan *Minkowski distance*.

##### 2. Mendefinisikan nilai K

Nilai  $k$  pada algoritma KNN mendefinisikan berapa banyak tetangga yang akan diperiksa untuk menentukan klasifikasi titik *kueri* tertentu. Misalnya, jika  $k=1$ , *instance* akan ditugaskan ke kelas yang sama dengan tetangga terdekatnya. Mendefinisikan  $k$  dapat menjadi tindakan penyeimbang karena nilai yang berbeda dapat menyebabkan *overfitting* atau *underfitting*. Nilai  $k$  yang lebih rendah dapat memiliki *varians* yang tinggi, tetapi bias yang rendah. Sedangkan

nilai k yang lebih besar dapat menyebabkan bias yang tinggi dan varians yang lebih rendah.

Pilihan k akan sangat bergantung pada data *input* karena data dengan lebih banyak *outlier* atau *noise* kemungkinan akan berkinerja lebih baik dengan nilai k yang lebih tinggi. Secara keseluruhan, disarankan untuk memilih nilai k berupa angka ganjil untuk menghindari ikatan dalam *klasifikasi*. Strategi *cross validation* juga dapat digunakan untuk membantu Anda memilih k yang *optimal* untuk *dataset* yang Anda miliki.

Langkah-langkah algoritma *KNN* yaitu:

1. Memasukkan data latih dan data uji
2. Menghitung nilai K
3. Menghitung jarak *Euclidian* setiap data latih terhadap data uji menggunakan

Rumus

$$(X, Y) = \sqrt{\sum_{i=1}^N (X_{i_{training}} - y_{i_{testing}})^2}$$

Keterangan :

$D(x,y)$ : jarak antara data latih dengan data uji,

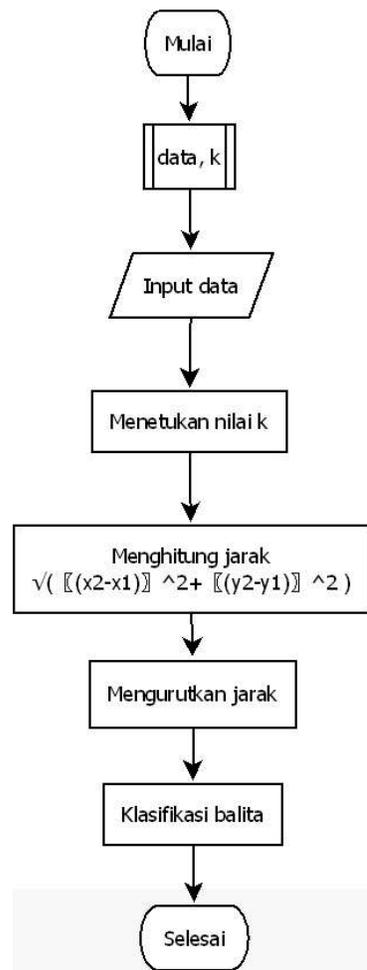
n : jumlah data latih

x : data latih

y : data uji

4. Mengurutkan hasil perhitungan jarak mulai dari yang terkecil ke yang terbesar.
5. Mengumpulkan atau mengambil sejumlah data sesuai nilai K yang telah ditentukan pada langkah
6. Menentukan hasil dari pengambilan data berdasarkan tetangga terdekat pada langkah 4 dapat diklasifikasikan berdasarkan kategori yang ditentukan.

### 2.2.5 Flowchart K-Nearest Neighbor



Gambar 2. 1 Flowcart KNN

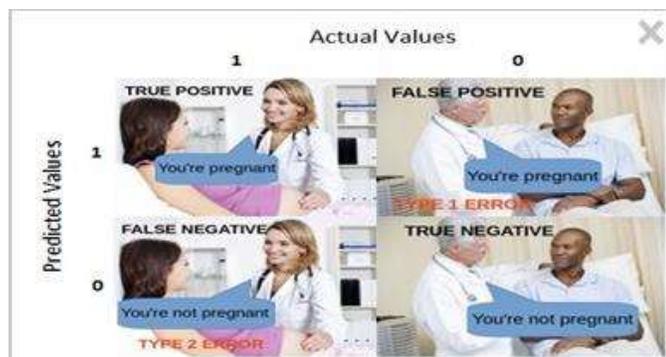
Dari *diagram flowcart* diatas dapat dijelaskan bahwa yang pertama kali dilakukan adalah mengumpulkan data bayi yang dibutuhkan seperti Nama, Jenis Kelamin, Berat badan, Tanggal imunisasi, dan Dosis. setelah terkumpul langkah berikutnya adalah menentukan nilai K. Nilai K merujuk pada jumlah tetangga terdekat yang akan digunakan untuk membuat prediksi.

*KNN* adalah algoritma pembelajaran mesin yang digunakan untuk *klasifikasi atau regresi* berdasarkan ketergantungan pada tetangga terdekat. Jika nilai K sudah diperoleh, maka langkah selanjutnya adalah menentukan jarak dan mengurutkan jarak. Dan tahap terakhir yaitu *Klasifikasi* data bayi sesuai dengan jumlah tetangga paling banyak.

#### **2.2.6 Confusion Matriks**

*Machine Learning* merupakan salah satu cabang dari disiplin ilmu kecerdasan buatan (*artificial intelligence*) yang membahas bagaimana sistem dibangun berdasarkan pada data. Jadi *machine learning* merupakan proses komputer untuk belajar dari data (*learn from data*). Jika tidak ada data, komputer tidak akan bisa belajar. Salah satu teknik aplikasi pada *machine learning* adalah *supervised learning*. Klasifikasi merupakan *supervised learning*, yang merupakan model prediksi dimana hasil prediksinya bersifat *diskrit*. Bagaimana mengukur *performa* dari model *klasifikasi* yang digunakan ? Jawaban sederhananya adalah membandingkan nilai aktual dengan nilai prediksi.

*Confusion Matrix* adalah pengukuran *performa* untuk masalah *klasifikasi machine learning* dimana keluaran dapat berupa dua kelas atau lebih. *Confusion Matrix* adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai *aktual*. Ada empat istilah yang merupakan *representasi* hasil proses *klasifikasi* pada *confusion matrix* yaitu *True Positif*, *True Negatif*, *False Positif*, dan *False Negatif*. Mari kita pahami apa itu *True Positif*, *False Positif*, *False Negatif*, dan *True Negatif* dalam *analogi* kehamilan (dapat dilihat pada Gambar dibawah ini).



Gambar 2. 2 Contoh *Confusion Matrix*

1. *True Positive (TP)*:

Interpretasi: memprediksi positif dan itu benar. Memprediksikan bahwa seorang wanita hamil dan wanita tsb memang benar hamil.

2. *True Negative (TN)*:

Interpretasi: Memprediksi negatif dan itu benar. Memprediksikan bahwa seorang pria tidak hamil dan benar ya pria kan tidak mungkin hamil

3. *False Positive (FP)*:

(Kesalahan Tipe 1)

Interpretasi: Memprediksi positif dan itu salah. Memprediksikan bahwa seorang pria hamil tetapi tidak mungkin pria bisa hamil

4. *False Negative (FN):*

(Kesalahan Tipe 2, kesalahan tipe 2 ini sangat berbahaya)

*Interpretasi:* Memprediksi negatif dan itu salah. Memperkirakan bahwa seorang wanita tidak hamil tetapi sebenarnya wanita tersebut hamil.

Dari contoh di atas dapat digambarkan bahwa:

1. Nilai Prediksi adalah keluaran dari program dimana nilainya *Positif dan Negatif*
2. Nilai Aktual adalah nilai sebenarnya dimana nilainya *True dan False*

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} S * 100\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} * 100\%$$

$$\text{Recall} = \frac{TP}{FN+TP} * 100\%$$

**Keterangan :**

- a. **Precision** *Presisi* dapat diartikan sebagai *derajat reliabilitas* model ketika ia memberikan prediksi "*Positif*".
- b. **Recall** *Recall* menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. Maka, *recall* merupakan *rasio* prediksi benar *positif* dibandingkan dengan keseluruhan data yang benar *positif*. Nilai *recall* dapat diperoleh dengan persamaan (3).

- c. *Accuracy* didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai *aktual*

### 2.2.7 Laravel

*Laravel* adalah sebuah *framework web* berbasis *PHP* yang *open-source* dan tidak berbayar, diciptakan oleh Taylor Otwell dan diperuntukkan untuk pengembangan aplikasi *web* yang menggunakan pola *MVC*. Struktur pola *MVC* pada *laravel* sedikit berbeda pada struktur pola *MVC* pada umumnya. Di *laravel* terdapat *routing* yang menjembatani antara *request* dari *user* dan *controller*. Jadi *controller* tidak langsung menerima *request* tersebut (Yudanto dkk, 2017).

### 2.2.8 Php MyAdmin

*phpMyAdmin* adalah perangkat lunak gratis (*freeware*) yang ditulis menggunakan bahasa pemrograman *PHP*, dengan dimaksudkan untuk menangani administrasi database *MySQL* melalui *interface Web*. *phpMyAdmin* mendukung berbagai operasi di database *MySQL* dan *MariaDB*. Operasi paling yang sering digunakan seperti mengelola *database*, tabel, kolom, *relasi*, *indeks*, pengguna, izin, dan lainnya, Dapat kita lakukan melalui antarmuka pengguna, sementara itu kita juga masih bisa menulis perintah *SQL* secara langsung untuk operasi pengelolaan *databasenya*.

### 2.2.9 MySql

*MySQL* adalah salah satu jenis *database server* yang sangat terkenal. *MySQL* menggunakan bahasa *SQL* untuk mengakses *database* nya. *Lisensi Mysql* adalah *FOSS License Exception* dan ada juga yang versi *komersial* nya. *Tag Mysql* adalah “*The World's most popular open source database*”. *MySQL* tersedia untuk

beberapa *platform*, di antaranya adalah untuk versi *windows* dan *versi linux*. Untuk melakukan administrasi secara lebih mudah terhadap *Mysql*, dapat menggunakan *software* tertentu, di antara nya adalah *phpmyadmin* dan *mysql*.

### **2.2.10 Php**

Menurut tim EMS(2021:61) *PHP* adalah Bahasa pelengkap *HTML* yang memungkinkan dibuatnya aplikasi *dinamis* yang memungkinkan adanya pengolahan data dan pemrosesan data. Semua *sintax* yang diberikan akan sepenuhnya dijalankan pada *server* sedangkan yang dikirimkan ke *browser* hanya hasilnya saja. Kemudian merupakan Bahasa berbentuk *script* yang ditempatkan didalam *server* dan diproses di *server*.

Hasilnya akan dikirimkan ke *client*, tempat pemakai menggunakan *browser*. *PHP* dikenal sebagai sebuah Bahasa *scripting*, yang menyatu dengan tag-tag *HTML*, dieksekusi di *server*, dan digunakan untuk membuat halaman *web* yang dinamis seperti halnya *Active Server Pages (ASP)* atau *Java Server Pages (JSP)*. *PHP* merupakan sebuah *software Open Source*