

## **BAB 2**

### **TINJAUAN PUSTAKA**

#### **2.1 Ginjal**

Pada manusia, ginjal merupakan organ utama yang melakukan proses ekskresi. Secara lengkap peranan atau fungsi ginjal adalah sebagai berikut [8]:

1. Mengekskresikan zat sisa seperti urea, asam urat, kreatinin, kreatin, dan zat lain yang bersifat racun;
2. Mengatur volume plasma darah dan jumlah air di dalam tubuh;
3. Menjaga tekanan osmosis dengan cara mengatur ekskresi garam – garam, yaitu membuang jumlah garam yang berlebihan dan menahan garam bila jumlahnya dalam tubuh berkurang;
4. Mengatur pH plasma dan cairan tubuh dengan mengekskresikan urin yang bersifat basa, tetapi dapat pula mengekskresikan urin yang bersifat asam;
5. Menjalankan fungsi sebagai hormon, dengan menghasilkan dua macam zat yaitu renin dan eritropoietin yang diduga memiliki fungsi endokrin.

Ginjal merupakan sepasang organ yang memiliki bentuk seperti kacang. Pada bagian luar ginjal disebut korteks, sedangkan bagian dalam ginjal disebut medula [8]. Pada bagian dalam terdapat ruang kosong yang disebut dengan pelvis. Dalam ginjal terdapat nefron yang merupakan unit struktural serta fungsional ginjal [8]. Nefron mengandung dua unsur, yaitu unsur pembuluh (elemen vaskuler) yang terdapat pada arterial, glomerulus (kumpulan kapiler), arterial eferan, dan kapiler tubuler dan unsur epitel yang terdapat pada kapsula Bowman, tubulus kontortus proksimal, lengkung henle, tubulus kontortus distal, dan tubulus kolektifus [8]. Anatomi ginjal dapat dilihat pada Gambar 2.1



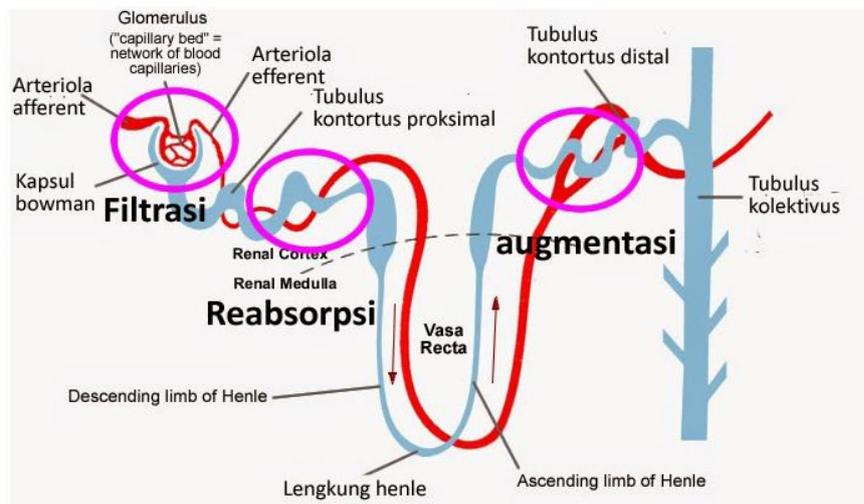
**Gambar 2.1 Struktur Ginjal**

Pada medula terdapat piramida ginjal dan piala ginjal yang mengandung pembuluh untuk mengumpulkan hasil ekskresi. Pembuluh tersebut langsung berhubungan dengan ureter yang bermuara di kantung kemih [8].

Nefron terdapat dua macam, yaitu nefron korteks yang terletak hampir seluruhnya pada korteks hanya sebagian kecil dari jerat henlenya yang berpenetrasi ke medula [8]. Nefron jukstamedular memiliki glomerulus [8] serta terletak pada perbatasan korteks – medula dan jerat henlenya masuk ke dalam medula [8].

### 2.1.1 Mekanisme Pembentukan Urin

Ginjal merupakan organ yang berfungsi sebagai sistem ekskresi yang merupakan proses pengeluaran zat sisa metabolisme dalam tubuh. Pada ginjal zat yang dikeluarkan berupa cairan urine. Proses pembentukan urin ini dapat dilihat pada Gambar 2.2 :



**Gambar 2.2 Proses Pembentukan Urin**

### 1. Filtrasi

Perpindahan cairan dari glomerulus menuju kapsula Bowman dengan menembus membran filtrasi yang terdiri dari tiga lapisan, yaitu sel endotelium glomerulus, membran basiler, dan epitel kapsula Bowman<sup>[1]</sup>. Hasil penyaringan berupa urin primer yang mengandung zat hampir sama dengan cairan yang menembus kapiler. Dalam keadaan normal urin primer tidak mengandung eritrosit, tetapi mengandung protein yang kadarnya kurang dari 0,03%, glukosa, garam, natrium, kalium, dan asam amino [8].

### 2. Reabsorpsi

Proses perpindahan cairan dari tubulus renalis menuju pembuluh darah yang mengelilinginya, yaitu kapiler peritubuler. Sel tubulus renalis secara selektif mereabsorpsi zat yang terdapat pada urin primer. Urin yang dihasilkan setelah reabsorpsi disebut urin sekunder (filtrat tubulus) [8]. Dalam keadaan normal, hampir 99% dari air yang menembus membran filtrasi akan direabsorpsi sebelum sampai ke ureter [8].

### 3. Augmentasi

Satu cara penambahan senyawa ke dalam filtrat (dari darah atau sel tubulus). Augmentasi merupakan proses aktif yang penting dalam menghilangkan obat – obatan, urea, dan kelebihan ion, serta menjaga keseimbangan asam – basa darah [8].

## 2.2 Penyakit Ginjal Kronis

Penyakit ginjal kronis (PGK) merupakan penyakit ginjal dimana ginjal memiliki *Glomerular filtration rate* (GFR) dibawah 60 mL/min/1.73m<sup>2</sup> selama lebih dari 3 bulan atau bukti terjadinya kerusakan ginjal selama lebih dari 3 bulan [2]. Penyakit ginjal kronis sering disebut “*silent disease*” karena penyakit ginjal kronis ini sulit untuk terdeteksi.

Di Indonesia penyakit ginjal kronis ini terdeteksi ketika laju filtrasi glomerulusnya sebesar 30% [1]. Hasil dari Riskesdas 2013, populasi umur  $\geq 15$  tahun yang terdiagnosis gagal ginjal kronis sebesar 0,2%. Perhimpunan Nefrologi Indonesia (Penaferi) pada tahun 2006 melakukan riset didapatkan bahwa prevalensi PGK sebesar 12,5%. Hasil Riskesdas 2013 menunjukkan bahwa angka prevalensi meningkat seiring dengan peningkatan pada umur yang meningkat tajam pada kelompok umur 35 – 44 tahun dibandingkan kelompok umur 25 – 34 tahun [1].

Pada ginjal kronis ini terdapat beberapa gejala – gejala awal yang akan dirasakan oleh penderita, seperti : tekanan darah tinggi, perubahan pada *urine*, adanya darah pada *urine*, kehilangan nafsu makan, mudah lelah, mual dan muntah, susah tidur, sakit kepala [1] [2]. Adapula faktor resiko terjadinya penyakit ginjal kronis, yaitu :

a. Hipertensi;

Menurut WHO hipertensi merupakan peningkatan atau tingginya tekanan darah. Di Indonesia, Hasil dari Riskesdas, prevalensi hipertensi pada penduduk dengan usia 18 tahun keatas adalah sebesar 25,8%. Namun penduduk yang terdiagnosis hipertensi oleh dokter hanya sebesar 9,4% [1].

b. Diabetes Melitus;

Diabetes melitus merupakan penyakit kronis yang terkait dengan kadar glukosa yang tinggi dalam darah [9]. Secara umum DM dikaitkan oleh obesitas, penuaan, penggunaan tembakau, aktivitas fisik, dan urbanisasi. Prevalensi diabetes melitus (DM) diseluruh dunia

diperkirakan akan sampai pada 366 juta pada tahun 2030, lebih dari 2 kali lipat pada tahun 2000.

Di Amerika Serikat penderita DM sudah mencapai 23,6 juta pasien dan 57 juta lainnya menderita pra-diabetes atau gangguan toleransi glukosa [9]. Menurut hasil Rikesdas 2013, di Indonesia prevalensi penderita diabetes adalah 5,7% dan hanya 26,3% yang sudah terdiagnosis [1].

c. Obesitas;

Obesitas merupakan faktor risiko kuat terjadinya PGK. Obesitas meningkatkan risiko faktor utama terjadinya PGK yaitu hipertensi dan diabetes melitus [1]. Pada pasien yang mengalami obesitas maka ginjal harus bekerja lebih keras untuk menyaring darah untuk memenuhi sistem metabolik akibat terjadinya peningkatan berat badan. Adanya penyaringan yang membutuhkan tenaga lebih ini dapat mengakibatkan ginjal mengalami kerusakan.

d. Riwayat penyakit anggota keluarga yang terdiagnosis PGK [2];

e. Umur  $\geq$  60 tahun.

Ketika penyakit ginjal kronis diduga, maka pasien harus melakukan serangkaian tes fungsi pada ginjal, yaitu [2]:

- a. Tes untuk albumin (tipe protein) dan/atau darah pada urin
- b. Tes darah untuk melihat tingkat produk sisa dalam darah dan menghitung laju filtrasi glomerulus.
- c. Tes tekanan darah jika penyakit ginjal ini disebabkan oleh tekanan darah tinggi dimana dapat merusak pembuluh darah pada ginjal.
- d. *CT scan* untuk mengambil gambar ginjal dan jejak urin. Tes ini untuk melihat ukuran ginjal serta letak tumor atau batu ginjal.

### 2.2.1 Tingkatan Penyakit Ginjal Kronis

Berdasarkan tingkat laju filtrasi glomerulus, fungsi ginjal dibagi menjadi 5 tingkatan yang dapat dilihat pada Tabel 2.1

**Tabel 2.1 Tingkat Penyakit Ginjal Kronis (PGK)**

	Tingkat	Kondisi Ginjal	Keterangan
--	---------	----------------	------------

<b>Early Stages</b>	1	eGFR normal $\geq$ 90mL/min/1,73m <sup>2</sup>	Sebagian orang tidak merasakan gejala PGK, namun penderita akan mengalami dehidrasi dan kepekaan yang tinggi terhadap obat-obatan
	2	Penurunan eGFR sedikit antara 60-89mL/min/1,73m <sup>2</sup>	
<b>Middle Stages</b>	3a	Penurunan eGFR ringan-sedang antara 45-59mL/min/1,73m <sup>2</sup>	Gejala PGK akan mulai terasa pada <i>stage</i> ini adalah adanya kandungan urea dan kreatinin dalam darah. Tekanan darah mulai naik serta munculnya gejala awal penyakit tulang dan anemia.
	3b	Penurunan eGFR sedang-berat antara 30-44mL/min/1,73m <sup>2</sup>	
	4	Penurunan eGFR berat antara 15-29 mL/min/1,73m <sup>2</sup>	
<b>End Stages</b>	5	Gagal ginjal, penurunan eGFR < 15m L/min/1,73m <sup>2</sup>	Perubahan pada frekuensi buang air kecil, hipertensi yang selalu datang, jumlah protein dalam urin meningkat, serta meningkatnya level kreatinin dan potasium dalam darah.

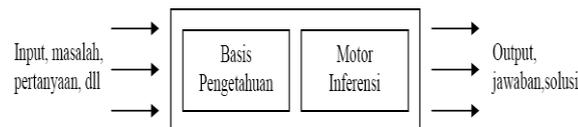
### 2.3 Artificial Intelligence

Menurut *Oxford Dictionary Artificial Intelligence* (AI) merupakan teori dan perkembangan sistem komputer untuk dapat melakukan pekerjaan yang biasanya membutuhkan kecerdasan manusia seperti, persepsi visual, pengenalan suara, pembuat keputusan, dan penerjemahan antar bahasa [10]. Tujuan dari AI ini adalah untuk membuat sistem yang ahli dimana sistem menunjukkan perilaku cerdas, belajar, menunjukkan, menjelaskan, dan memberi saran kepada penggunanya. Selain itu AI dibuat untuk mengimplementasikan kecerdasan manusia kedalam mesin dimana menciptakan sistem yang memahami, berpikir, belajar, dan berperilaku seperti manusia [11]. Adapun pengertian *artificial intelligence* menurut para ahli, yaitu [12] :

1. H.A.Simon (1987) “kecerdasan buatan (*artificial intelligence*) merupakan kawasan penelitian, aplikasi, dan intruksi yang terkait dengan pemrograman komputer untuk melakukan hal yang dalam pandangan manusia adalah cerdas”

2. Rich dan knight (1991) mengartikan “kecerdasan buatan (*artificial intelligence*) sebagai sebuah studi tentang bagaimana membuat komputer melakukan hal – hal yang pada saat ini dapat dilakukan dengan baik oleh manusia”

Terdapat dua buah bagian utama yang dibutuhkan dalam membentuk kecerdasan buatan [12] yang dapat dilihat pada Gambar 2.3



**Gambar 2.3 Bagian Utama AI**

1. Basis Pengetahuan

Kecerdasan buatan perlu memiliki basis pengetahuan yang berisi fakta-fakta, teori, pemikiran, dan hubungan antara satu dengan yang lainnya.

2. Motor Inferensi

Kecerdasan buatan perlu memiliki motor inferensi yang merupakan kemampuan mesin untuk menarik kesimpulan berdasarkan pengalaman yang ada.

## 2.4 *Machine Learning*

*Machine Learning* atau biasa disebut pembelajaran otomatis merupakan cabang dari *Artificial Intelligence* (AI) dimana program komputer dapat “belajar” dari masukan yang tersedia. Pembelajaran adalah proses mengubah pengalaman menjadi keahlian atau pengetahuan. Masukan kedalam algoritma adalah data pelatihan yang mewakili pengalaman dan keluaran merupakan keahlian [12]. *Machine learning* membantu dalam mencari solusi terhadap berbagai permasalahan dalam visualisasi, pengenalan suara, dan robotika [13].

*Machine Learning* telah bercabang menjadi beberapa sub bidang yang berhubungan dengan berbagai jenis pembelajaran. Dalam *machine learning* pembelajaran dibagi menjadi 3, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning* [12].

### 2.4.1 *Supervised Learning*

*Supervised learning* merupakan pembelajaran dimana tujuannya adalah untuk mempelajari pemetaan dari masukan ke keluaran yang nilai – nilainya yang benar sudah disediakan [13]. Pada *supervised learning* diperlukan adanya data *training* untuk dilakukan pelatihan serta data *testing* untuk pengujian. Pada *supervised learning* ini dibagi menjadi 2 jenis, yaitu :

- a. Regresi merupakan pembelajaran pada *machine learning* dimana melakukan pemetaan dari variabel masukan ke variabel keluaran [13]. Variabel keluaran pada regresi ini berupa nilai nyata atau kontinu. Contoh pada *supervised learning* regresi ini adalah memprediksi umur seseorang dilihat dari gambar wajahnya.
- b. Klasifikasi merupakan pembelajaran pada *machine learning* dimana melakukan pemetaan dari variabel masukan ke variabel keluaran [13]. Variabel keluaran pada klasifikasi biasa disebut dengan label atau kategori. Contoh pada *supervised learning* klasifikasi ini adalah memprediksi tumor pada pasien apakah tumor itu berjenis *benign* atau *malignant*.

### 2.4.2 *Unsupervised Learning*

*Unsupervised learning* merupakan pembelajaran dimana tujuannya untuk menemukan keteraturan dalam masukan [13]. Adanya struktur ke ruang *input* sehingga terbentuk pola – pola tertentu. Pada *unsupervised learning* metode yang digunakannya yaitu *clustering* yang bertujuan untuk menemukan kelompok atau pengelompokkan masukan.

## 2.5 *Imbalance Data*

Salah satu permasalahan utama pada *data mining* adalah ketidakseimbangan kelas. *Dataset* dikatakan tidak seimbang jika satu kelas (dikatakan kelas mayoritas) jauh lebih banyak jumlahnya dibanding kelas yang lain (dikatakan kelas minoritas). Salah satu masalah yang muncul dari ketidakseimbangan *dataset* adalah keefektifan akurasi (tingkat kesalahan) serta menentukan kinerja suatu

metode pengklasifikasian [14]. Dari permasalahan itu, maka salah satu cara untuk mengatasi ketidakseimbangan kelas adalah metode *sampling*.

Sampling merupakan metodologi populer untuk mengatasi ketidakseimbangan kelas. Tujuan dari metode pengambilan sampel adalah untuk membuat kumpulan data memiliki distribusi kelas yang seimbang [14]. Ada dua metode, yaitu *under-sampling technique* dan *over-sampling technique*.

### **2.5.1 *under-sampling technique***

Teknik ini dilakukan dengan cara membuang sejumlah data yang ada pada kelas mayoritas hingga keseimbangan data dicapai [14]. Salah satu kelemahan utama pada teknik ini adalah informasi yang berguna akan hilang ketika dilakukan pembuangan data secara acak [14].

### **2.5.2 *over-sampling technique***

Teknik ini dilakukan dengan cara melakukan penyalinan dan pengulangan data yang ada pada kelas minoritas sehingga keseimbangan data tercapai. Salah satu kelemahan pada teknik ini adalah terjadinya masalah *overfitting*. Untuk mengatasi masalah tersebut salah satu teknik *over-sampling* yang lain adalah dengan membuat data sintesis pada kelas minoritas atau biasa disebut *synthetic minority over-sampling technique (SMOTE)* [14].

## **2.6 Metode *Waterfall***

Model *waterfall* pertama kali dikenalkan pada tahun 1970 oleh Winston Royce. *Waterfall* membagi pembangunan perangkat lunak ke dalam beberapa tahap yang dilakukan secara terurut [15]. Model *waterfall* menggambarkan proses pengembangan perangkat lunak dalam aliran sekuensial linier. Ini berarti bahwa setiap tahapan dalam proses pengembangan dimulai hanya jika tahapan sebelumnya selesai. Dalam model *waterfall* ini, tahapan tidak tumpang tindih [16].

Penggunaan model *waterfall* ini dapat dilakukan ketika teknologi dapat dipahami dan tidak bersifat dinamis, pengerjaannya singkat, serta kebutuhan di dokumentasikan secara baik, jelas dan pasti [16].

### 2.6.1 Keuntungan *Waterfall*

Berikut ini merupakan keuntungan dari metode *waterfall* [17] :

1. Meyediakan tahap demi tahap pemeriksaan untuk pembangunan perangkat lunak
2. Hanya perlu melanjutkan ke tahap berikutnya setelah tahap sebelumnya telah selesai dikerjakan
3. *Waterfall* dapat diaplikasikan pada sebuah pendekatan yang berulang – ulang.

### 2.6.2 Kekurangan *Waterfall*

Selain adanya keuntungan pada *waterfall*, adapula kekurangan pada metode *waterfall* ini. Berikut merupakan kekurangan pada metode *waterfall* [17]

1. Adanya umpan balik minimal diantara tahap pembangunan perangkat lunak
2. Pada *waterfall* akan mulai melihat hasilnya nanti di *life cycle*
3. Setiap tahapan dilacakan dengan terlalu banyak hari dan *milestones*.

## 2.7 Probabilitas

Probabilitas merupakan kemungkinan kemunculan suatu peristiwa [18]. Teori probabilitas ada hubungannya dengan eksperimen dengan hasil yang berbeda. Contoh eksperimen tersebut termasuk pelemparan koin dua sisi yang memiliki dua hasil yaitu kepala dan ekor, memilih seseorang dari suatu populasi dan menentukan apakah orang tersebut perokok dengan dua hasil perokok atau non-perokok [19]. Ketika eksperimen sudah terdefinisi dengan baik, maka kumpulan semua hasil disebut dengan ruang sampel [19]. Secara matematis, ruang sampel merupakan himpunan sedangkan hasilnya disebut dengan elemen dari himpunan [19]. Pada ruang sampel terbatas, setiap ruang sampel disebut dengan peristiwa dan bagian ruang sampel yang memiliki satu elemen disebut dengan peristiwa dasar [19]. Adapun Aturan peluang untuk kejadian A,B, dan  $\emptyset$  adalah sebagai berikut; [20]

2.  $0 \leq P(A) \leq 1$

Peluang suatu kejadian nilainya terletak diantara 0 dan 1 inklusif.

3.  $P(\emptyset) = 0$   
Peluang himpunan kosong adalah 0
4.  $P(A) \leq 1 - P(A)$   
Peluang komplemen suatu kejadian adalah 1 – peluang kejadian tersebut
5.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

### 2.7.1 Peluang Bersyarat

Peluang bersyarat banyak digunakan dalam dunia bisnis dan ekonomi. Salah satu contohnya adalah penggunaan teorema Bayes yang banyak digunakan dalam pengambilan keputusan [21]. Kejadian bersyarat dikenal pula dengan kejadian yang saling bergantung. Dua kejadian, misalnya A dan B adalah kejadian bersyarat jika kejadian B terjadi karena kejadian A terjadi sehingga kejadian A mempengaruhi munculnya kejadian B [22]. Peluang bersyarat A bila B diketahui dengan  $P(A|B)$  dan didefinisikan sebagai [23]

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) P(B|A)}{P(B)}; P(B) \neq 0 \quad (2.1)$$

dengan,  $P(A|B)$  = peluang kejadian bersyarat atau peluang kejadian A jika B sudah terjadi,  $P(A \cap B)$  = peluang kejadian saling bebas (dua buah kejadian yang tidak saling mempengaruhi), dan  $P(B)$  = peluang munculnya kejadian yang dimaksud dari kejadian pertama,  $P(B|A)$  = peluang kejadian bersyarat atau peluang kejadian B jika A sudah terjadi.

Teorema atau aturan Bayes yang dikenal dengan teorema mengenai probabilitas sebab – sebab memungkinkan untuk menentukan probabilitas dari berbagai kejadian  $A_1, A_2, \dots, A_n$  yang dapat menyebabkan A terjadi. Misalnya  $A_1, A_2, \dots, A_n$  adalah kejadian – kejadian saling meniadakan yang gabungannya merupakan ruang sampel S, dengan kata lain, salah satu dari kejadian – kejadian tersebut harus terjadi. Maka jika A adalah kejadian sembarang, maka aturan Bayes dapat didefinisikan sebagai berikut;

$$P(A_k|A) = \frac{P(A_k) P(A|A_k)}{\sum_{j=1}^n P(A_j) P(A|A_j)} \quad (2.2)$$

## 2.8 Bayesian Networks

*Bayesian Networks* atau biasa disebut sebagai *Belief Networks* atau *probabilistic causal networks* merupakan grafik struktural untuk mewakili relasi probabilitas antara sebagian besar variabel dengan melakukan inferensi probabilistik terhadap variabel – variabel tersebut [19].

*Bayesian networks* merupakan metode klasifikasi probabilistik dimana setiap fiturnya memiliki relasi sehingga dibutuhkan pemodelan untuk melihat relasi antar fiturnya. Pemodelan pada *bayesian networks* adalah *directed acyclic graph* (DAG) dimana *node* dalam *bayesian networks* digambarkan sebagai *random variable* sedangkan *arch* merupakan relasi langsung antar fitur atau variabel. Model pada DAG dibagi menjadi 4, yaitu [19] :

- a.  $X \rightarrow Z \rightarrow Y$  (*head to tail*) : *arch / edge* bertemu di kepala ke ekor Z, Z merupakan simpul *head to tail* pada rantai.
- b.  $X \leftarrow Z \rightarrow Y$  (*tail to tail*) : *arch / edge* bertemu di ekor ke ekor Z, Z merupakan simpul *tail to tail* pada rantai.
- c.  $X \rightarrow Z \leftarrow Y$  (*head to head*) : *arch / edge* bertemu di kepala ke kepala Z, Z merupakan simpul *head to head* pada rantai.
- d.  $X - Z - Y$  dimana X dan Y tidak berdekatan, simpul ini merupakan pertemuan yang tidak dapat dipisahkan.

Pada *bayesian networks* terdapat dua tahap pembelajaran yaitu, *structural learning* dan *parameter learning* :

### 2.8.1 Structural Learning

*Structural learning* merupakan tahap untuk mengidentifikasi topologi pada *bayesian networks* [24]. Langkah – langkah umum yang dilakukan dalam membuat *structural learning*, yaitu :

2. Pilih ordering variabel ( $X_1, \dots, X_n$ )
3. Lakukan perulangan dari 1 sampai n
  - a. Tambahkan  $X_i$  ke dalam networks
  - b. Pilih Parent dari  $X_1, \dots, X_{i-1}$  sehingga  $P(X_i | \text{Parent}(X_i)) = P(X_i | X_1, \dots, X_{i-1})$

4. Agar *bayesian networks* sah, maka  $X_i$  harus *conditionally independent* terhadap semua  $X_1, \dots, \dots, X_{i-1}$  yang bukan anggota dari *parent*  $X_i$

Pada pembangunan *structural learning* terdapat tiga pendekatan, yaitu:

a. *Search-and-Score based methods*

Pada pendekatan *search and score* ini memperlakukan *structural learning* sebagai masalah pemilihan model. pada tahap *score* ini dibuat fungsi *score* untuk mengevaluasi struktur yang sebelumnya dibuat dari data *training*. Pada tahap *search* untuk melakukan pencarian pada nilai *score* yang tertinggi [25].

b. *Constrain-based approaches*

pada pendekatan *constraint based* variabel diuji dengan cara menghitung independensi bersyarat untuk menemukan dependensi dan independensi antar variabelnya. Setelah itu jaringan *bayesian* akan dibangun untuk dilihat independensinya [25].

c. *Bayesian Model averaging*

Pada pendekatan *model averaging* ini membuat ensemble dengan menggunakan metode *structural learning* yang ada yaitu *constraint-based* dan *search-and-score based* [25]. Pada pendekatan ini alih – alih menggunakan salah satu model dan melakukan inferensi. *Model averaging* melakukan inferensi dengan cara menggunakan masing – masing pola DAG lalu mengalikan hasil probabilitasnya dengan kemungkinan posterior dari struktur [19].

### 2.8.2 *Parameter Learning*

*Parameter learning* merupakan tahap pembelajaran dalam *bayesian networks* dimana pada tahap ini melakukan estimasi pada parameter numerik atau probabilitas bersyarat sesuai dengan topologi jaringan yang ada [24]. Pada *bayesian networks* DAG disebut sebagai struktur sedangkan nilai – nilai pada fitur disebut sebagai parameter [19].

Pada tahap pembelajaran di *parameter learning* ini dilakukan dengan cara menghitung *conditional probability distribution* [19]. Pada kasus dimana grafik sudah diketahui, maka dapat menggunakan perhitungan sebagai berikut :

a. MLE (*Maximum Likelihood Estimation*)

MLE merupakan metode yang sering digunakan untuk mengestimasi nilai parameter dari data yang modelnya sudah diketahui. Suatu nilai parameter dikatakan bagus dilihat dari seberapa banyak kemungkinan untuk menghasilkan data yang diamati [26]. Perhitungan MLE ini dilihat dari value yang dimiliki pada setiap fitur apakah fitur memiliki binomial atau multinomial value. Pada fitur yang merupakan binomial value dapat dihitung dengan menggunakan rumus [26]

$$\theta = \frac{Na}{Na + Nb} \quad (2.3)$$

Pada fitur yang memiliki multinomial value, maka estimasi value dapat dihitung menggunakan rumus

$$\theta = \frac{Nk}{\sum_{\lambda} N_{\lambda}} \quad (2.4)$$

b. MAP (*Maximum a Posteriori Estimation*)

$$P(\mu|X) = \frac{P(X|\mu)P(\mu)}{P(X)} \quad (2.5)$$

Hukum Bayes mengubah parameter  $\mu$  menjadi probabilitas posterior ( $P(\mu|X)$ ). Dengan kemungkinan fungsi  $P(X|\mu)$ , maka estimasi posterior (MAP) didefinisikan sebagai [27]

$$\mu_{MAP} = \underset{\mu}{arg \max} P(\mu|X) \quad (2.6)$$

Sedangkan perhitungan nilai parameter dimana grafik belum diketahui, data yang memiliki nilai *value* yang berbeda – beda, maka perhitungan estimasi setiap parameternya dengan menggunakan *dirichlet distribution*. *Dirichlet Distribution* memiliki *density* yang dapat dihitung dengan rumus [28] :

$$Dir_{a_1, a_2, \dots, a_n}(x_1, \dots, x_n) := \frac{\Gamma(a_1 + \dots + a_n)}{\Gamma_{a_1} \dots \Gamma_{a_n}} x_1^{a_1-1} \dots x_{n-1}^{a_{n-1}-1} \quad (2.7)$$

### 2.8.3 Inferensi *Bayesian Networks*

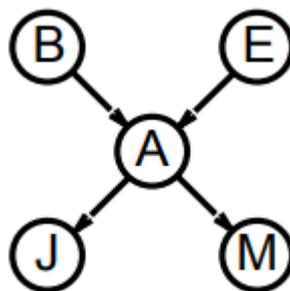
Inferensi pada *bayesian networks* merupakan cara pengambilan keputusan dalam metode *bayesian networks*. Pada metode *bayesian networks* cara umum untuk menentukan keputusan dengan menggunakan rumus [29]

$$\operatorname{argmax} P(y|x) = \prod_{u \in U} P(u|pa(u)) \quad (2.8)$$

Selain dapat menggunakan inferensi yang ada secara umum, *bayesian networks* memiliki dua pendekatan pengambilan keputusan untuk menentukan klasifikasi, yaitu :

a. Enumerasi

Enumerasi merupakan salah satu inferensi yang ada pada *bayesian networks* dimana pengambilan keputusannya dilakukan dengan cara menjumlahkan semua parameter yang ada sesuai dengan struktur yang telah dibuat sebelumnya [30]. Perhitungan inferensi dengan enumerasi ini dapat dilihat pada kasus alarm dimana strukturnya dapat dilihat pada Gambar 2.4



**Gambar 2.4 Struktur Kasus Alarm**

Diberikan contoh bagaimana untuk mengetahui peluang *burglary* (B) ada jika john(J) dan marry menelepon(M). Maka pengambilan keputusan dari contoh kasus yang ada dengan menggunakan enumerasi adalah sebagai berikut

$$P(B|J, M) = P(B) \sum_e P(e) \sum_a P(a|b, e) P(J|a) P(M|a) \quad (2.9)$$

b. Eliminasi Variabel

Selain dapat dilakukan dengan enumerasi pengambilan keputusan dapat menggunakan eliminasi variabel. Eliminasi variabel dilakukan

dengan cara melakukan penjumlahan dari kiri ke kanan, lalu menyimpan hasil pemfaktoranannya untuk menghindari rekompulasi [30]. Dengan melihat contoh kasus pada Gambar 2.3, maka pengambilan keputusan dengan menggunakan eliminasi variabelnya adalah [31]

$$\begin{aligned}
 & P(B|J, M) \\
 &= P(B) \sum_e P(e) \sum_a P(a|b, e) P(J|a) P(M|a) \\
 &= P(B) \sum_e P(e) \sum_a P(a|b, e) P(J|a) f_m(a) \\
 &= P(B) \sum_e P(e) \sum_a f_a(a, b, e) f_j(a) f_m(a) \\
 &= P(B) \sum_e P(e) f_{ajm}(b, e) \text{ (jumlahkan terhadap } a) \\
 &= P(B) f_{ejm}(b) \text{ (jumlahkan terhadap } e) \\
 &= f_b(b) \times f_{ejm}(b)
 \end{aligned} \tag{2.10}$$

## 2.9 *k-fold cross validation*

*Cross validation* merupakan teknik validasi model untuk menilai bagaimana hasil analisis statistik akan menggenerasikan ke kumpulan data independen [32]. *Cross validation* digunakan untuk melihat seberapa akurat kinerja model yang dibuat. Selain itu *cross validation* dapat digunakan untuk melakukan komparasi performa pada berbagai metode klasifikasi yang berbeda [32]. *Cross validation* memiliki beberapa metode untuk melakukan validasi, salah satunya adalah *k-fold cross validation*.

*k-fold cross validation* merupakan teknik yang biasa digunakan untuk mengestimasi kesalahan saat pengujian dimana estimasi ini dapat digunakan untuk melihat model terbaik dan memberikan gagasan tentang kesalahan pengujian model yang dipilih paling terakhir [32]. Metode yang digunakan untuk menghitung *k-fold cross validation* ini, yaitu [33] :

1. Bagi data kedalam bagian K yang hampir sama.
2. Untuk setiap  $k = 1, 2, \dots, K$ , cocokkan model dengan parameter  $\lambda$  ke dalam bagian  $k-1$ , diberikan  $\beta^{-k}(\lambda)$  dan menghitung *error* dalam memprediksi bagian  $k$  dengan rumus

$$E_k = \sum_{i \in kth\ part} (y_i - x_i \beta^{-k}(\lambda))^2 \quad (2.11)$$

3. Untuk menghitung *cross validation error* menggunakan rumus

$$CV = \frac{1}{K} \sum_{k=1}^K E_k(\lambda) \quad (2.12)$$

