

## BAB IV

### METODOLOGI PENELITIAN

#### 4.1 Data

Data pada penelitian tersebut diambil berdasarkan data *website* <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10072> yang merupakan suatu *server* yang memuat informasi bioteknologi yang menyangkut DNA, protein, senyawa aktif dan taksonomi. *Platform* data yang digunakan adalah GSE10072 data ini adalah *gene expression* data kanker paru-paru *non small cell lung cancer*

#### 4.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian merupakan *gene expression data* adapun variabel yang digunakan adalah :

##### 1. *Gen*

Gen yang digunakan berasal dari *gene expression* data kanker paru-paru *non small cell lung cancer* (NSCLC)

##### 2. Jaringan

Jaringan merupakan terdiri dari kelompok sel yang struktural dengan fungsional yang sama. Terdapat dua jaringan pada data diantaranya :

###### a. *Adenocarcinoma of the Lung*

Merupakan sejenis kanker yang terbentuk pada jaringan kelenjar yang mengeluarkan sekresi lendir di seluruh tubuh dan dapat terjadi pada paru-paru.

###### b. *Normal Lung Tissue*

Merupakan kanker yang terjadi pada jaringan normal paru-paru.

##### 3. Jenis Kelamin

Variabel jenis kelamin pada data digunakan untuk membedakan penderita kanker paru memiliki jenis laki-laki atau perempuan berdasarkan status merokok. Ozlu dan Bulbul (2005) menyatakan penderita kanker paru-paru 90% laki-laki dan 79% adalah perempuan memiliki status sebagai perokok.

#### 4. Status Merokok

Status merokok merupakan faktor resiko terbesar pada kanker paru-paru menurut Ozlu dan Bubbul (2005) resiko seseorang perokok terkena kanker paru-paru akan meningkat sebesar 20-40 kali. Status merokok dibagi menjadi tiga yaitu masih merokok sampai saat ini, tidak pernah merokok, dan pernah merokok.

#### 5. Stadium

*Stadium* adalah pengelompokan dari stadium yang memberikan tanda bahwa kanker yang diderita merupakan masih pada tahap awal atau sudah tahap lanjutan dimana pada stadium I (satu) hingga II (dua) akan digolongkan sebagai stadium awal atau tahap awal sedangkan pada III (tiga) hingga IV (empat) di golongkan sebagai stadium akhir.

#### 4.3 Metode Analisis Data

Metode yang digunakan adalah K-NN untuk mengklasifikasikan *gene expression data* dengan menggunakan bantuan *software* Rstudio v10.153

#### 4.4 Alat dan Cara Organisasi Data

Alat yang digunakan adalah *software* Rstudio vs10.153 dan cara organisasi data dengan cara mendownload data yang berasal *website* NCBI dengan link <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10072> data yang didapatkan dalam bentuk .CEL yang akan diolah *software* melalui beberapa proses sehingga didapatkan *gene expression data*

