BAB I
PENDAHULUAN

1.1 Latar Belakang

Bioinformatika adalah suatu bidang ilmu yang memiliki peranan penting dalam manajemen data informasi genetika. Bioinformatika dapat digunakan untuk mengidentifikasi, mendiagnosa dan pengobatan suatu penyakit. Perkembangan bioinformatika di Indonesia pada bidang biologi hanya digunakan oleh peneliti biologi melokul, hal tersebut disebabkan oleh keterbatasan dalam menggunakan alat dan analisis data. Bidang TI sekalipun bioinformatika masih kurang mendapat perhatian(Aprijani dan Elfaizi, 2004).

Salah satu penyakit yang dapat di diagnosis dengan menggunakan penerapan bioinformatika adalah kanker. Kanker adalah suatu penyakit yang disebabkan oleh pertumbuhan dan penyebaran yang tidak terkontrol dari seltidak normal (ACS,2017). Kanker memiliki faktor resiko yang dapat meningkat dalam tubuh melalui paparan karsinogen, konsumsi alkohol, serta penggunaan rokok. Kanker adalah penyebab kematian terbesar didunia, terhitung pada tahun 2015 terdapat 8,8 juta kematian yang dikarenakan kanker dan tercatat 1,68 juta lainya adalah kaker paru-paru (WHO,2017). Penderita kanker paru-paru tercatat bahwa 50–60% penderita merupakan penderita *adenocarcinoma* pada *non- small cell lung cancer* (Klamerus dkk, 2009).

Stadium atau stadium pada kanker paru-paru dapat di bagi menjadi stadium I (satu) hingga stadium IV (empat). Pengukuran stadium pada kanker dapat dilakukan dengan melihat ukuran tumor,*lymph node*, dan terjadinya metasis (AJCC,2017). Stadium kanker sangat berpengaruh dalam penentuan tindakan untuk pengobatan kanker.

Metode klasifikasi yang digunakan adalah*K-nearest neighbor* (K-NN) yang merupakan salah satu metode klasifikasi *unsupervised mechine learning.*K-NN biasanya juga disebut non-parametrik klasifikasi karena tidak adanya asumsi yang

mengikuti metode tersebut. K-NN akanmengklasifikasikan suatu pengamatan baru kedalam kelas yang sama dengan suatu pengamatan dari *training set* yang paling dekat dengan suatu pengamatan baru. Metode ini sangat sederhana digunakan dan cukup baik.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, sehingga disusunlah rumusan masalah sabagai berikut:

1. Bagaimana data *microarray gene expression* di proses ?
2. Bagaimana menentukan nilai *k* terbaik dari metode K-NN untuk mengklasifikasikan stadium pasien kanker paru-paru?
3. Bagaimana hasil dari klasifikasi stadium pasien kanker paru-paru berdasarkan gene *expression data* menggunakan K-NN?
4. Bagaimana tingkat akurasi dari klasifikasi stadium pasien kanker paru-paru berdasarkan gene *expression data* menggunakan K-NN ?

1.3 Tujuan Penulisan

Berdasarkan rumusan masalah diatas tujuan dari penelitian memahami dan mengetahui cara memproses dan mengklasifikasikan dengan menggunakan K-NN pada *microarray gene expression data*

1.4 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut :

1. Manfaat bagi penulis adalah pengaplikasian ilmu statistik yang telah didapatkan.
2. Memberikan gambaran hasil klasifikasi stadium kanker peru-paru berdasarkan*gene expression data*, GSE10072 dengan menggunakan metode K-NN .

1.5 Batasan Masalah

Penelitian memiliki batasan yaitu sebagai berikut :

1. Data yang digunakan adalah *Microarray gene expression data* GSE10072 yang merupakan data pasien kanker paru-paru dengan jumlah sampel 107.
2. Metode statistik yang digunakan adalah *K-nearest neighbor* (K-NN).
3. *Software* statsistika yang digunakan adalah Rstudio v10.1