

## **BAB II**

### **TINJAUAN PUSTAKA**

Landi dkk (2008) dalam “*Gene Expression Signature of Cigarette Smoking And Its Role in Lung Adenocarcinoma Development And Survival* ” yang menyatakan bahwa dilakukan analisis *gene expression* dengan menggunakan *chip* Affymetrix HG-U133A pada 135 jaringan sampel adenokarsinoma paru-paru dengan status perokok, tidak pernah merokok dan pernah merokok dan di validasi dengan menggunakan teknik biokimia. Analisis ANOVA digunakan untuk pembauran potensi, pengujian prosedur, *gene set enrichment analysis*, dan klasifikasi GO digunakan untuk pemilihan gen. Hasil diidentifikasi *gene expression* status merokok yang mencangkup siklus gen. *Gene expression* membedakan antara perokok dan tidak perokok di paru-paru hingga pada tahap awal tumor. Didapatkannya pula bahwa ekspresi NEK2 ( $p = 0.001$ ) dan TTK ( $p=0.002$ ) meningkat 3 kali resiko kematian akibat adenokarsinoma paru pada perokok.

Liu dkk (2016) dalam “*Identification of Feature Genes For Smoking-Related Lung Adenocarcinoma Based on Gene Expression Profile Data*” yang menjelaskan bahwa dalam penelitian bertujuan untuk mengidentifikasi jalur gen yang terkait dengan hubungan merokok pada adenocarcinoma paru. Tiga dataset yang digunakan adalah GSE43458, GSE10072, dan GSE50081. Subjek yang digunakan adalah perokok dan tidak perokok, berdasarkan *feature gene* akan dibangun jaringan interaksi protein dan pemilihan gen akan di optimalkan dengan menggunakan metode CC. Hasil dari analisis jalur didapatkan total 83 gen tergeneralisasi yang terekspresikan secara regional dan dibangun jaringan PPI. Kemudian dilakukan klasifikasi dengan menggunakan metode SVM secara akurat mengidentifikasi 27 *feature gene* sehingga didapatkan hasil klasifikasi 89 perokok dari 93 diklasifikasikan dengan benar dan 19 dari 23 bukan perokok diklasifikasikan dengan benar dengan nilai akurasi klasifikasi 93.10%

Almoeirfi (2015) dalam "*Classification Approaches For Microarray Gene Expression Data*" yang memaparkan bahwa klasifikasi biologis merupakan salah satu analisis yang paling banyak dilakukan pada data *microarray* yang bertujuan untuk mengklasifikasikan gen yang tidak diketahui berdasarkan data dengan menggunakan beberapa metode klasifikasi yang nantinya akan dibandingkan kebaikan model di antara metode tersebut adalah SVM, K-NN, dan NN, dengan menggunakan empat kumpulan dataset *microarray* antara lain adalah dataset leukimia, prostat, usus besar, dan payudara. Analisis akan memvariansikan persentase distribusi dataset ditambah dengan dataset *training set* dan *testing set* untuk memastikan didapatkan hasil terbaik.

Dawan (2008) dalam "*Application of Committee K-NN Classifiers For Gene Expression Profile Classification*" yang memaparkan bahwa penelitian yang dilakukan bertujuan untuk merancang suatu sistem klasifikasi yang stabil dan mengkategorikan *microarray gene expression profile*. *High-throughput microarray* telah digunakan untuk melihat *gene expression* dalam sampel biologis. Data yang digunakan dalam penelitian adalah data leukimia mitoxantrone. Tahap *preprocessing* sekitar 38% gen dikeluarkan, kemudian gen yang informatif akan disingkat dan digunakan untuk pengelompokan K-NN. Hasil analisis didapatkan bahwa K-NN memberikan hasil terbaik. Hasil K-NN akan dievaluasi dengan menggunakan komite K-NN dan didapatkan pula evaluasi memberikan akurasi dan stabilitas yang lebih baik dari K-NN.

Maher dkk (2014) dalam "*Classification of Two Types Cancer Based on Microarray Data*" yang menyatakan bahwa kemajuan teknologi dalam bidang biologi meningkatkan dan teknik pembelajaran *machine learning* dengan klasifikasi dataset kanker (*microarray expression dataset*) dibagi menjadi dua langkah untuk mencapai hasil akurasi yang diinginkan. Langkah pertama adalah menerapkan teknik seleksi gen untuk mengurangi jumlah gen yang terlibat dalam proses klasifikasi, mengurangi waktu dan meningkatkan efisiensi. Langkah kedua adalah menyesuaikan kekuatan klasifikasi untuk mencapai hasil akurasi yang kuat untuk

sampel baru yang tidak terklasifikasi. Peneliti mempresentasikan bahwa *t-test* mampu menghasilkan ranking pada gen dengan baik dan mampu untuk mengurangi data. K-NN juga menjadi klasifikasi yang cukup menjajikan dengan akurasi proporsi untuk mengklasifikasikan gen yang tidak ketahui berdasarkan data *microarray*. Penelitian digunakan beberapa metode klasifikasi yang nantinya akan dibandingkan kebaikan modelnya diantara metode tersebut adalah SVM, K-NN, dan NN, dengan menggunakan empat kumpulan dataset *microarray*. Dataset yang digunakan adalah leukimia, prostat, usus besar, dan payudara. Analisis akan memvariansikan persentase distribusi dataset ditambah dengan dataset *training set* dan *testing set* untuk memastikan didapatkan hasil terbaik. Hasil penelitian menjelaskan bahwa klasifikasi SVM memberikan hasil akurasi yang lebih baik dibandingkan klasifikasi lainnya.