

# Model integrasi data teks dan gene expression untuk tujuan prediksi ikatan obat dan penyakit pada kanker menggunakan bayesian network = Integration data model of text and gene expression to prediction of drug and disease association on cancer using bayesian network

Edo Surya Utama, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20404196&lokasi=lokal>

---

## Abstrak

Reposisi obat merupakan proses penting yang digunakan untuk meningkatkan proses penemuan obat, prosesnya dilakukan dengan memanfaatkan banyaknya sumber data dan tipe data seperti data dari dokumen dan data gene expression. Metode yang digunakan untuk mengolah dokumen adalah Text Mining, metode ini mengekstraksi data menjadi sebuah informasi yang berguna. Data gene expression adalah data hasil hibridisasi terhadap sekuen nukleotida dan dari keseluruhan data gene expression tersebut dipilih satu kelompok kecil gen yang merupakan gen yang aktif berkembang jika seseorang mengidap penyakit kanker. Metode pengelompokan yang diusulkan adalah metode Gene Shaving yang dapat mengidentifikasi himpunan bagian dari data gene expression dengan pola ekspresi yang koheren dan varian tinggi. Salah satu permasalahan di dalam mengolah data dengan tipe data yang berbeda adalah pengintegrasian seluruh data. Pendekatan yang diusulkan pada penelitian ini untuk integrasi data adalah Bayesian Network. Tujuan dari integrasi data adalah untuk memprediksi ikatan antara obat dan penyakit. Hasil prediksi ikatan obat dan penyakit yang didapatkan dari model integrasi data teks dan gene expression adalah 81,69%. Hasil ini meningkat dibandingkan dengan hasil prediksi ikatan obat dan penyakit dengan hanya menggunakan satu tipe data saja, di mana nilai prediksi ikatan obat dan penyakit dengan menggunakan tipe data teks adalah 70,58% dan nilai prediksi ikatan obat dan penyakit menggunakan tipe data gene expression adalah 66%.

<hr>

Drug repositioning is an important process that is used to enhance the drug discovery process, the process is carried out by using the number of data sources and types of data such as data from gene expression data and documents. Text Mining is one of method to process of text, this method of data extraction into a useful information. Gene expression is the data from the hybridization of the nucleotide sequence and gene expression of the whole data set is selected a small group of genes that are active genes evolve if someone with cancer. Clustering method that proposed is Gene Shaving, that can identify subsets of genes with expression patterns are coherent and high variants. One of the problems in process of data with different data types is integration all of data. The approach proposed in this study for data integration is Bayesian Netwotk. The purpose of data integration is to predict the bond between the drug and disease. The result of drug and disease prediction bond obtained from the integration model was 81,69 %. These results increased compared with the predicted results by using one type of data, PPV`s values prediction of association drug and disease by text data type was 70,58 % whereas PPV`s values prediction of association drug and disease by gene expression data type was 66%.