Abstrak

Cancer atau yang dikenal secara umum oleh masyarakat Indonesia sebagai penyakit tumor ganas merupakan penyakit yang diakibatkan dari pertumbuhan tidak normal dalam sel jaringan tubuh. Penyakit kanker termasuk dalam kategori penyakit yang berbahaya yang dapat menyebabkan kematian. Berdasarkan data dari Badan Kesehatan Dunia – WHO, tahun 2015 tidak kurang dari 8,8 juta manusia meninggal karena penyakit kanker. Jumlah tersebut kian meningkat setiap tahun. Walaupun begitu, 30 hingga 50% dari penyakit ini dapat dicegah.

Oleh karena itu penelitian ini bertujuan untuk mengurangi tingkat kematian. Dengan mengumpulkan data pasien yang berupa informasi tentang pasien yang mengidap penyakit kanker merupakan langkah awal dalam menekan tingkat kematian. *Microarray* DNA (ekspresi gen) telah banyak digunakan dalam penelitian kanker untuk memprediksi hasil penyakit.

Penelitian yang dilakukan menggunakan metode *Minimum Redundancy Maximum Relevance* (mRMR) untuk menyeleksi jumlah fitur, dan metode *Probabilistic Neural Network* (PNN) untuk mengklasifikasikan. Dengan metode seleksi fitur yang digunakan, diperoleh 10 fitur terbaik. Dari 10 fitur tersebut dapat di klasifikasikan jenis penyakit yang diderita pasien berdasarkan probabilitas jarak antar fitur. Akurasi yang didapatkan dari algoritma klasifikasi PNN sebesar 90.91%. Data ekspresi gen yang digunakan adalah data *leukemia* (kanker darah) diperoleh dari *Kent Ridge Biomedical Data set Repository*.

Kata kunci : data mining, klasifikasi, minimum redundancy maximum relevance, probabilistic neural network