

ABSTRACT

Tawang Wulandari, Cancer Classification Based on Protein Structure Using Naïve Bayes Algorithm

Supervisor: Drs. Marji, M.T. and Lailil Muflikhah S.Kom, M.Sc.

Cancer begins when there are abnormal cells that divide uncontrollably. Cancer can be formed due to the transformation of a gene, Due to the continuously splitting nature of the cell that is supposed to form new cells to replace the old cells, otherwise make the abnormal cells continue to divide and continue to grow. Every year there are thousands of new cases of cancer affecting Indonesians. The late detection caused many deaths from cancer. It is predicted that the number of cancer patients the next 20 years will reach 24 million patients.

The exact cause of cancer is not known for certain. External factors causing cancer include genetic, carcinogen and lifestyle. These factors can alter the structure of DNA. DNA changes are detrimental to the process of cell division and beneficial to the mutation process. In the mutation process can produce p53 gene, the genetic changes are most commonly found in human cancers. From these problems required a system to classify the types of cancer suffered by patients. One of the methods used is the Naive Bayes algorithm. Naive Bayes is a simple probability classification that applies Bayes Theorem with the assumption of high dependence. The algorithm is known to have been widely used in the field of medicine. The algorithm is applied to matters relating to medical diagnosis. Diagnosis is done by looking at the related symptoms then looking at the probability of the possibility of the disease.

In this thesis is tested using several datasets, namely 320, 400, 480, 588 and 848 datasets. The percentage of test data taken from the dataset is 10% to 60%, respectively. Tests conducted to determine the level of accuracy affected by the percentage of test data and the number of protein structure datasets. Accuracy results obtained in the test 848 dataset with the percentage of 60% test data obtained the highest accuracy of 79,17%.

Keywords: cancer, protein structure, naive bayes