

Abstract

HIV (Human Immunodeficiency Virus) is a virus that damages the immune system. TB is a bacterial infection disease. The reason, cannot rely on symptoms to know whether the body has been infected with HIV. People who are infected with HIV will be more susceptible to other diseases such as tuberculosis, for example. To detect TB in people with HIV usually by means of a tuberculin skin test, sputum sampling, and chest X-rays. To detect tuberculosis in HIV patients, microarray technology can also be used. Microarray is a technology to look at the sequence of nucleic acid sequences in certain parts to analyze the structure of thousands of gene expressions that are present in certain parts of the body simultaneously. Microarray data on TB cases in HIV patients has very large dimensions. Therefore, machine learning methods are needed to facilitate analysis. In this final project, the implementation of machine learning (machine learning) on HIV patients is carried out using the ensemble method, namely Random Forest. This method includes classification that produces accuracy values and F1-scores. In the results of this final project, Random Forest has three features to find the best value, namely run 10, run 20, and run 30 and the best accuracy is 0.833 and 0.866 with f-1 scores of 0.823, 0.799, and 0.875.

Keywords: HIV; tuberculosis; microarrays; ensemble method; machine learning; Random Forest.