

ABSTRAK

Nama : Dhea Juliani
Program Studi : Teknik Informatika
Judul : Implementasi *Machine Learning* untuk Klasifikasi Penyakit Kanker Paru Menggunakan Metode *Naïve Bayes* dengan Tambahan Fitur *Chatbot*.
Dosen Pembimbing : Muhamad Soleh, S.Si, M.Kom

Menurut *The Global Burden of Cancer* diperkirakan kasus baru kanker meningkat menjadi 19,3 juta pada tahun 2020, dengan sekitar 10 juta kematian. Kanker paru merupakan jenis kanker terbanyak di Indonesia, dengan sekitar 34 ribu kasus baru dan 30 ribu kematian pada tahun 2020. Salah satu bentuk kemajuan teknologi informasi adalah sistem deteksi penyakit kanker paru dan sistem *chatbot*, sistem *chatbot* digunakan untuk mempermudah pengguna dalam berkomunikasi dan mendapatkan informasi mengenai penyakit kanker paru. Tahapan yang akan dilakukan adalah pengumpulan *dataset*, *text preprocessing*, perancangan tampilan antarmuka sistem deteksi dan *chatbot*, dan klasifikasi *naïve bayes*. *Dataset* penyakit kanker paru diperoleh dari *website Kaggle* dengan jumlah 309 data dan 16 fitur diantaranya *Gender*, *Age*, *Smoking*, *Yellow Fingers*, *Anxiety*, *Peer Pressure*, *Chronic Disease*, *Fatigue*, *Allergy*, *Wheezing*, *Alcohol Consuming*, *Coughing*, *Shortness of Breath*, *Swallowing Difficulty*, dan *Chest Pain*. *Dataset* pertanyaan untuk *chatbot* merupakan data yang dimasukkan secara manual dan disimpan dalam format JSON dan untuk pemodelan pada sistem ini menggunakan *Artificial Neural Network* (ANN) dengan pendekatan *Natural Language Processing* (NLP). Hasil klasifikasi *naïve bayes* memperoleh tingkat akurasi tertinggi yang dicapai dalam penelitian ini sebesar 81%, tetapi untuk sistem *chatbot* dan sistem deteksi penyakit kanker paru masih belum terintegrasi.

Kata Kunci : Kanker paru, *Machine Learning*, *Chatbot*, *Naïve Bayes Classifier*.

ABSTRACT

According to The Global Burden of Cancer, it is estimated that new cases of cancer will increase to 19.3 million by 2020, with around 10 million deaths. Lung cancer is the most common type of cancer in Indonesia, with around 34 thousand new cases and 30 thousand deaths in 2020. One form of information technology advancement is a lung cancer disease detection system and chatbot system, the chatbot system is used to facilitate users in communicating and getting information about lung cancer. The stages that will be carried out are dataset collection, text preprocessing, interface design of the detection system and chatbot, and naïve bayes classification. The lung cancer dataset is obtained from the Kaggle website with 309 data and 16 features including Gender, Age, Smoking, Yellow Fingers, Anxiety, Peer Pressure, Chronic Disease, Fatigue, Allergy, Wheezing, Alcohol Consuming, Coughing, Shortness of Breath, Swallowing Difficulty, and Chest Pain. The question dataset for the chatbot is data that is manually entered and stored in JSON format and for modeling in this system using an Artificial Neural Network (ANN) with a Natural Language Processing (NLP) approach. The naïve bayes classification results obtained the highest accuracy rate achieved in this study of 81%, but the chatbot system and the lung cancer detection system are still not integrated.

Keywords : Lung Cancer, Machine Learning, Chatbot, Naïve Bayes Classifier.