

ABSTRAK

Kanker serviks merupakan penyebab kematian nomor dua pada perempuan di dunia setelah kanker payudara, sedangkan di Indonesia kanker serviks menduduki peringkat pertama, hal tersebut yang menjadikan masalah kesehatan reproduksi di Indonesia masih menjadi sorotan utama. *K-Nearest Neighbor* (KNN) adalah suatu metode algoritma *supervised learning*, di mana kelas yang paling banyak muncul (mayoritas) yang akan menjadi kelas hasil klasifikasi. Sedangkan Algoritma C4.5 merupakan sebuah algoritma klasifikasi yang digunakan untuk membangun *decision tree* (pohon keputusan). Penelitian kali ini bertujuan untuk mengklasifikasikan data resiko kebiasaan kanker serviks dengan menerapkan algoritma KNN dan C4.5. Data diambil dari *website UCI Machine Learning* sebanyak 72 data dan 19 atribut setelah dilakukan seleksi menjadi 63 data dan 5 atribut yang diantaranya adalah dukungan sosial instrumental, pengetahuan pemberdayaan, kemampuan pemberdayaan dan keinginan pemberdayaan lalu untuk kanker serviks dijadikan untuk atribut kelas. Pengujian ini dilakukan dengan cara manual, pemrograman *python* dan *rapidminer*. Penghitungan algoritma KNN telah dilakukan pada pengujian menggunakan *rapidminer* dengan *cross validation* kemudian menghasilkan akurasi 80.95% dan dengan *split validation* membagi data menjadi data *training* dan *testing* sebesar 80 : 20 menghasilkan akurasi 83.33%, sedangkan algoritma C4.5 dengan *cross validation* kemudian menghasilkan akurasi 76.19% dan dengan *split validation* membagi data menjadi data *training* dan *testing* sebesar 80 : 20 menghasilkan akurasi 75.00%. Untuk pengujian dengan pemrograman *python* dengan *split validation* membagi data menjadi data *training* dan *testing* sebesar 80 : 20 kemudian algoritma KNN mendapatkan hasil akurasi 84.00%, sedangkan algoritma C4.5 menghasilkan akurasi 69.00%. Sehingga algoritma KNN dengan pengujian *Python* mendapatkan akurasi terbaik pada penelitian ini dengan nilai akurasi 84.00%.

Kata Kunci : Klasifikasi, Kanker Serviks, *K-Nearest Neighbor* (KNN), C4.5

ABSTRACT

Cervical cancer became the second leading cause of death in women in the world after breast cancer, while in Indonesia cervical cancer has the highest position, which causes reproductive health problems in Indonesia still a major priority. K-Nearest Neighbor (KNN) is a supervised learning algorithm method, in which the class that appears the most (the majority) will be the class resulting from the classification. While the C4.5 algorithm is a classification algorithm used to build a decision tree (decision tree). This study aims to classify the risk data for cervical cancer habits by applying the KNN and C4.5 algorithms. Data were taken from the UCI Machine Learning website as many as 72 data and 19 attributes after being selected to become 63 data and 5 attributes including instrumental social support, empowerment knowledge, empowerment ability and desire for empowerment and then cervical cancer was used as a class attribute. This test is done manually, programming python and rapidminer. The calculation of the KNN algorithm has been carried out on testing using rapidminer with cross-validation then producing an accuracy of 80.95% and with split validation dividing the data into training and testing data of 80: 20 resulting in an accuracy of 83.33%, while the C4.5 algorithm with cross validation then produces an accuracy of 76.19% and with split validation dividing the data into training and testing data of 80: 20 resulting in an accuracy of 75.00%. For testing with python programming with split validation dividing the data into training and testing data of 80: 20 then the KNN algorithm gets 84.00% accuracy results, while the C4.5 algorithm produces 69.00% accuracy. So that the KNN algorithm with Python testing gets the best accuracy in this study with an accuracy value of 84.00%.

Keywords: Classification, Cervical Cancer, K-Nearest Neighbor (KNN), C4.5