International Journal Information System and Computer Science (IJISCS)

Full Paper
eISSN : 2598-246X
pISSN : 2598-0793

SERVICES CANCER DETECTION SYSTEM USING K-NEAREST NEIGHBOURS(K-NN) METHOD AND NAÏVE BAYES CLASSIFIER

*Corresponding author

Sri Rezeki Candra Nursari[1], Nanda Mahya Barokatun Nisa[2]
[1,2] Pancasila University, Faculty of Engineering, Information Technology, Jakarta - Indonesia

sri.rezeki.candra.n@univpancasila.ac.id
nandaamahyaa@gmail.com

Abstract
According to WHO (World Health Organization) data, every 2 minutes a woman dies. In Indonesia alone, 40 - 45 women are diagnosed with cervical cancer every day. Of those diagnosed, around 20-25 die from cervical cancer. About 95% more cervical cancer is caused by infection with the HPV virus or the human papilloma virus and an estimated death rate reaches 270,000 deaths each year. Cervical cancer occupies the third rank type of cancer in the world after breast and lung cancer, because the symptoms are not very visible at an early stage, so it is referred to as "Silent Killer". Based on the data and cases above, the latest technology that is able to detect cervical cancer in order to speed up the detection process for someone to be quickly treated is an artificial intelligence application that serves to detect whether someone should run 4 cervical cancer testing techniques, namely Hinselmann, Schiller, Citology, and biopsy with K-nearest neighbors algorithm and Naive Bayes classifier is one of the latest technologies that can facilitate the work of a doctor and speed up the process of detecting someone whether to run 4 testing techniques or not. The correct amount of data classified by the K-Nearest Neighbors method is 558 data from 858 data. The classification accuracy of the Naïve Bayes method is 84.7%. The correct amount of data classified by the Naïve Bayes method is 558 data from 858 data. The classification accuracy of the Naïve Bayes method is 84%.
Keywords: K-Nearest Neighbours, KNN, Naïve Bayes Classifier, Kanker Serviks

1.0 INTRODUCTION

Diagnosis of diseases using a computer has many advantages, so the results of computer processing can be used as a second opinion for doctors and low cost. In addition, the computer-based diagnosis can be a solution to improve the accuracy of diagnosis without going through medical procedures. Cervical cancer is one type of cancer that appears on a woman's cervix, the cervix itself serves as an entrance to the uterus from the vagina. The main cause of cervical cancer is the Human Papilloma virus (HPV).

According to WHO (World Health Organization) data, every 2 minutes a woman dies. In Indonesia alone, 40 - 45 women are diagnosed with cervical cancer every day. Of those diagnosed, around 20-25 die from cervical cancer. About 95% more cervical cancer is caused by infection with the HPV virus or the human papilloma virus and an estimated death rate reaches 270,000 deaths each year.

2.0 METHODOLOGY

2.1. K-Nearest Neighbours(K-NN)

K-Nearest Neighbors Algorithm (K-NN) is a method that uses a supervised algorithm. K-NN is included in the group based learning. This algorithm is also one of the lazy learning techniques. K-NN is done by finding the k group of objects in the training data closest to the object in the new data or testing data. In general, to define the distance between two objects x and y, the Euclidean distance formula is used.

$$dist\ (x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Information :
$dist$ (x, y) = euclidean distance
xi = training data
yi = data testing
n = amount of research data

K-Nearest Neighbors algorithm sequence:

1. Determine the parameter K (number of closest neighbors)
2. Calculate the euclidean distance of each object against the given sample data.
3. Sort the objects into groups that have the smallest euclid distance.
4. Collecting the category Y (classification of earest neighbors).
5. By using the majority category, it can be predicted the value of the instance query that has been calculated.

2.2. Naïve Bayes Classifier

Bayes classifier is one of the statistical classifiers, where this classifier can predict the probability of class membership of a tuple data that will enter into a particular class, according to probability calculations. Bayes classifiers are based on the Bayes theorem discovered by Thomas Bayes in the 18th century. In a comparative study of classification algorithms we have found a simple bayesian or commonly known as the Naïve Bayes classifier. Naïve Bayes classifier shows high accuracy and speed when applied to large databases.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Information :
X = Data with unknown classes
H = Data hypothesis is a specific class
P (H | X) = H hypothesis probability based on condition X (posteriori probability)
P (H) = Hypothesis probability H (prior probability)
P (X | H) = Probability of X based on conditions on the hypothesis H
P (X) = Probability of X
Naive Bayes algorithm sequence:

1. Calculate P (Ci) for each class
2. Calculate P (X | Ci) for each criterion and each class
1. Look for the largest P (X | Ci) being the conclusion

4.0 RESULANTS AND DISCUSSION

The results and discussion of cervical cancer detection systems using the K-Nearest Neighbors (K-NN) Method and Naïve Bayes Classifier are

1. The data obtained from UCI data have many missing values, therefore the data that have missing values use data replacement methods using mean, median, mode.
2. The results of calculations using these 2 methods, the K-Nearest Neighbors (K-NN) accuracy is superior to the Naïve Bayes Classifier. You can see the accuracy value of the 2 methods below:

Table 1. Accuracy

Accuracy	*K-Nearest Neighbors(KNN)*	*Naïve Bayes Classifier*
Hinselmann	0.847176 (84,7%)	0.840531 (84%)
Schiller	0.847176 (84,7%)	0.840531 (84%)
Citology	0.847176 (84,7%)	0.840531 (84%)
Biopsi	0.847176 (84,7%)	0.840531 (84%)

5.0 CONCLUSION

Based on the results of the study, that the K-Nearest Neighbors (KNN) and Naïve Bayes Classifier methods are applied equally for training and testing data, namely 35% Tesing, 65% Training with Random 40. Then the results of KNN with a distance of 5 with an accuracy rate of

84.7%, Naive Bayes of 4 targets with an accuracy rate of 84%. It can be concluded from these 2 methods that the K-Nearest Neighbors is better than Naïve Bayes Calssifier.

REFERENCES

[1] Agung Wibowo. (2017). "Aplikasi Diagnosis Penyakit Kanker Payudara Menggunakan Algoritma Sequential Minimal Optimization". Jurnal Teknologi dan Sistem Komputer, 5(4), 2017, 153-158.

[2] Nurul Baety Tsani, Harliana, dan Abdul Rachman. (2019), "Implementasi Deteksi Tepi Canny Dengan Transformasi Powerlaw Dalam Mendeteksi Stadium Kanker Serviks". Jurnal Ilmiah INTECH (Information Technology Journal) of UMUS Vol.01, No.01, Mei 2019, pp. 22~33

[3] Siti Rochwati, Sutopo Patria Jati, dan Antono Suryoputro. (2016. "Pengetahuan Bidan Mempengaruhi Praktik Bidan dalam Konseling Pemeriksaan IVA pada Wanita Usia Subur". Jurnal Promosi Kesehatan Indonesia Vol. 11 No. 2 Agustus 2016

[4] Kementrian Kesehatan Republik Indonseia. (2018). "Data dan Informasi Profil Kesehatan Indonesia". Jakarta : Kementerian Kesehatan RI

[5] Ella Nurlelawati, Tria Eni Rafika Devi, dan Imas Sumiati. (2018). "FAKTOR-FAKTOR YANG BERHUBUNGAN DENGAN KEJADIAN KANKER SERVIKS DI RUMAH SAKIT PUSAT PERTAMINA JAKARTA AHUN 2016". Jurnal Bidan "Midwife Journal" Volume 5 No. 01

[6] Yunita Marliana. (2014). "AKURASI METODE INSPEKSI VISUAL DENGAN ASAM ASETAT/IVA UNTUK DETEKSI DINI KANKER LEHER RAHIM". Jurnal Kesehatan Prima Volume : 8, No.2, Agustus 2014, Hal : 1336-1344

[7] Nonik Ayu Wantini, Novi Indrayani. (2019). "Deteksi Dini Kanker Serviks dengan Inspeksi Visual sam Asetat (IVA)". JURNAL NERS DAN KEBIDANAN DOI:10.26699/jnk.v6i1.ART.p027–034

[8] Nanda Aulia Rizqi dan Novri Suhermi. "Predicting Cervical Cancer Risk Classification". Institut Teknologi Sepuluh Nopember (ITS) Surabaya

[9] Tutus Praningki dan Indra Budi. (2017). " Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN". Citec Journal, Vol. 4, No. 2, Februari 2017 – April 2017 ISSN: 2460-4259

[10] Rahayu Prihatini Saputri. "Penerapan Data Mining Technique sebagai Evaluasi Ketepatan Akurasi terhadap Klasifikasi Mushroom Data Set". Institut Teknologi Sepuluh Nopember

[11] Silfia Andini. (2013). "KLASIFIKASI DOKUMENT TEKS MENGGUNAKAN ALGORITMA NAÏVE AYES DENGAN BAHASA PEMOGRAMAN JAVA". JURNAL TEKNOLOGI INFORMASI & PENDIDIKAN VOL. 6 NO. 2 ISSN : 2086 – 4981

[12] MEI LESTARI. (2014). "PENERAPAN ALGORITMA KLASIFIKASI NEAREST NEIGHBOR (K-NN) UNTUK MENDETEKSI PENYAKIT JANTUNG". Faktor Exacta 7(4): 366-371, 2014 ISSN: 1979-276X

[13] Miftahil Fauza, Aprianti, dan Azrimaidaliza. (2019). "Faktor yang Berhubungan dengan Deteksi Dini Kanker Serviks Metode IVA di Puskesmas Kota Padang". Jurnal Promosi Kesehatan Indonesia Vol. 14 No. 1

[14] Fitri Handayani dan Feddy Setio Pribadi. (2015). "Implementasi Algoritma Naiv e Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110". Jurnal Teknik Elektro Vol. 7 No. 1

[15] Y. M. S. Al-Wesabi, Avishek Choudhury, Daehan Won. (2018). "Classification of Cervical Cancer Dataset". Proceedings of the 2018 IISE Annual Conference

[16] Riyan Latifahul Hasanah dkk. (2019). "KLASIFIKASI PENERIMA DANA BANTUAN DESA MENGGUNAKAN METODE KNN (K-NEAREST NEIGHBOR)". Jurnal TECHNO Nusa Mandiri Vol. 16, No. 1

[17] Ermatita. (2016). "ANALISIS DAN PERANCANGAN SISTEM INFORMASI PERPUSTAKAAN". Jurnal Sistem Informasi (JSI), VOL. 8, NO. 1, April 2016, ISSN : 2355-4614

[18] Awalia Ardiana, dan Amak Yunus EP. "SISTEM PREDIKSI PENENTUAN JENIS TANAMAN SAYURAN BERDASARKAN KONDISI MUSIM DENGAN PENDEKATAN METODE TREND MOMENT". Teknik Informatika, Universitas Kanjuruhan Malang

[19] Hermawan. "PREDIKSI PENERIMAAN SISWA BARU PADA MADRASAH ALIYAH AS-SAYAFI'IYAH 02 MENGGUNAKAN METODE TIME SERIES". Jurnal Ilmu Komputer dan Sistem Informasi
[20] WEN WU dan HAO ZHOU. (2017). "Data-Driven Diagnosis of Cervical Cancer With Support Vector Machine-Based Approaches". IEEEAcccess VOLUME 5, 2017
[21] DWI AYU MAHARDIKA. "MANAJEMEN KAMPANYE PROGRAM KONMUTER (KONSUMEN MUDAH TERLAYANI) UNTUK MENINGKATKAN PELAYANAN PUBLIK". Studi Deskriptif Kualitatif tentang Manajemen Kampanye Program Konmuter pada PT. PLN PerseroDistribusi Jawa Timur
[22] Chesley Tanujaya. (2017). "PERANCANGAN STANDART OPERATIONAL PROCEDURE PRODUKSI PADA PERUSAHAAN COFFEEIN". Jurnal Manajemen dan Start-Up Bisnis Volume 2, Nomor 1, April 2017
[23] Yuli Asriningtias dan Rodhyah Mardhiyah. (2014). "APLIKASI DATA MINING UNTUK MENAMPILKAN INFORMASI TINGKAT KELULUSAN MAHASISWA". JURNAL INFORMATIKA Vol. 8, No. 1
[24] Anandary Riezka, Imelda Atastina dan Kiki Maulana. "ANALISIS DAN IMPLEMENTASI DATA-CLEANING DENGAN MENGGUNAKAN ETODE MULTI-PASS NEIGHBORHOOD (MPN)". Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom
[25] Abidatul Izzah dan Nur Hayatin. (2013). "Imputasi Missing data Menggunakan Algoritma Pengelompokan Data K-Harmonic Means". Seminar Nasional Matematika dan Aplikasinya 2013
[26] Kadek Wibowo. (2015). "ANALISA KONSEP OBJECT ORIENTED PROGRAMMING PADA BAHASA PEMROGRAMAN PHP". JURNAL KHATULISTIWA INFORMATIKA, VOL. 3, NO. 2
[27] Jan Wira Gotama Putra. (2019). "PENGENALAN KONSEP PEMBELAJARAN MESIN DAN DEEP LEARNING". Tokyo. Jepang
[28] Purbo, Onno W. (2019). "Text Mining – Analisis Medsos, Kekuatan Brand, & Intelijen di Internet". ANDI (Anggota IKAPI)
[29] Prabowo Pudjo Widodo, Rahmadya Trias Handayanto, Herlawati. (2013). "PENERAPAN DTA MINING DENGAN MATLAB". Rekayasa Sains, Bandung.
[30] Budiharto, Widodo. (2018). "PEMROGRAMAN PYTHON UNTUK ILMU KOMPUTER DAN TEKNIK". ANDI (Anggota IKAPI)
[31] Mustakim dan Giantika Oktaviani F. (2016). "Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa". Jurnal Sains, Teknologi dan Industri, Vol. 13, No.2 ISSN 2407-0939
[32] Effrida Manalu, Fricles Ariwisanto Sianturi dan Mamed Rofendy Manalu. (2017). "PENERAPAN ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI JUMLAH PRODUKSI BARANG BERDASARKAN DATA PERSEDIAAN DAN JUMLAH PEMESANAN PADA CV. PAPADAN MAMA PASTRIES". Jurnal Mantik Penusa Vol. 1 No.2