

Instance Selection dengan Naïve Bayes pada Klasifikasi Kanker Serviks

Fari Katul Fikriah

Teknologi Informasi/Institut Teknologi dan Sains Nahdlatul Ulama Pekalongan

email: farichatulfikriyah45@gmail.com

DOI: <https://doi.org/10.31603/komtika.v5i2.6041>

Received: 13-10-2021, Revised: 02-11-2021, Accepted: 26-11-2021

ABSTRACT

There are several deadly disease for woman, one of which is cervical cancer. The growth and development of the disease is very slow, so that treatment if know form the beginning will facilitate the healing process, but conversely unknown cancers from the beginning will become dangerous and deadly disease due to relatively difficult healing. Biopsy action is one way to detect the presence of cancer. In the previous study, classification of cervical cancer had the biggest accuracy value of 97,515% using the decision tree method of several feature selection technique. for this reason, this research will use the decision tree or tree C4.5 classification method, logistic function and zero R which were previously carried out processing with instance selection with Naïve Bayes by reducing the elimination of missing values with the aim of increasing the level of accuracy better than previous studies. C4.5 classification in this study has the most maximum results compared to other classification methods with an accuracy value of 99,69%.

Keywords: *Classification, Cervical Cancer, Instance Selection, Naïve Bayes, C4.5.*

ABSTRAK

Terdapat beberapa penyakit mematikan bagi wanita salah satunya adalah kanker serviks. Tumbuh kembang penyakit tersebut sangatlah lambat, sehingga pengobatan tersebut jika diketahui dari awal akan mempermudah proses penyembuhan, namun sebaliknya kanker yang tidak diketahui dari awal akan menjadi penyakit berbahaya dan mematikan karena penyembuhan yang relatif sulit. Tindakan biopsi menjadi salah satu cara untuk mendeteksi adanya penyakit kanker. Pada penelitian sebelumnya klasifikasi terhadap kanker serviks memiliki nilai akurasi paling tinggi sebesar 97.515% dengan menggunakan metode *Decision Tree* dari teknik pemilihan beberapa fitur. Untuk itu penelitian ini akan menggunakan metode klasifikasi *Decision Tree* atau *Tree C4.5*, *Logistic Function* dan *ZeroR* yang sebelumnya dilakukan *preprocessing* dengan *instance selection* dengan *Naïve Bayes* dengan mengurangi menghilangkan *missing value* dengan tujuan menaikkan tingkat akurasi yang lebih baik dari penelitian sebelumnya. Klasifikasi *C4.5* dalam penelitian ini memiliki hasil paling maksimal dibandingkan dengan metode klasifikasi lain dengan nilai akurasi sebesar 99,69%.

Kata-kata kunci: *Klasifikasi, Kanker Serviks, Instance Selection, Naïve Bayes, C4.5.*

PENDAHULUAN

Kebiasaan hidup sehat saat ini harus selalu diterapkan oleh siapa saja, karena tanpa manusia sadari pola hidup sehari-hari menjadi penyebab seseorang terjangkit oleh suatu penyakit tertentu. Pola kehidupan yang sehat adalah kebiasaan dimana hidup harus berpegang pada prinsip dalam menjaga kesehatan, maka tidak sedikit orang yang melanggar prinsip tersebut mudah untuk terpapar suatu penyakit dan salah satunya adalah penyakit kanker serviks. Kanker serviks merupakan salah satu penyakit mematikan sekaligus ganas yang lambat laun dapat menurunkan daya tahan tubuh penderitanya, pola hidup yang tidak baik serta kurang menjaga kebersihan pada daerah intim kewanitaan dapat menjadi penyebab terdiagnosanya penyakit tersebut. Kelompok sel yang tidak normal pada mulut rahim wanita

dapat menimbulkan kanker serviks. Penyakit tersebut umum terjadi dikalangan perempuan dan merupakan salah satu penyakit yang dapat dicegah. Tumbuh biak kanker serviks ditandai dengan perubahan tubuh seseorang dan relatif berkembang sangat lambat. Sel skuamusa dan glandulas merupakan sel yang mempengaruhi adanya kanker serviks, namun sebagian besar dipengaruhi oleh karnisoma sel skuamosa [1]. Penyakit kanker serviks akan membahayakan nyawa penderita jika terjadi keterlambatan dalam vonis kanker serviks itu sendiri. Kanker itu sendiri merupakan salah satu kanker ginekologi tersering yang menyerang wanita, sekaligus menjadi salah satu penyakit kanker yang dapat dideteksi dan dapat diobati sepenuhnya pada tahap prakanker. Meskipun pada umumnya banyak penderita kanker serviks yang berakhir meninggal dunia [2] [3].

Dalam ilmu teknologi masih terdapat identifikasi kanker serviks yang memiliki akurasi rendah. Pada penelitian yang dikemukakan oleh beberapa peneliti salah satunya Sofjayana, et.al [4]. Tujuan penelitian tersebut adalah mengidentifikasi stadium kanker dengan membandingkan data kanker menggunakan metode J48. Akurasi yang didapatkan pada penelitian tersebut adalah sebesar 93,03% dengan hasil sensitifitas diatas 0,8 pada semua tahap data kanker. Sunny Sharma [5] pada penelitiannya mengidentifikasi stadium kanker serviks dengan metode klasifikasi menggunakan *Decision Tree* menghasilkan akurasi 67,5%.

Pada penelitian sebelumnya menggunakan model klasifikasi *tree C45* untuk menentukan tingkat akurasi pada penentuan tindakan biopsi pada penyakit kanker serviks dengan kesimpulan bahwa akurasi yang didapat bisa lebih dari 97,5%. Penelitian tersebut dilakukan oleh Y.M.S. A-Wesabi dan A. Choudhury [6]. Penelitian yang dilakukan oleh Kelwin Fernandes, et al [7] metode skrining atau diagnosis dan preferensi dokter subjektif menciptakan ekosistem yang kompleks untuk metode otomatis pada pasien kanker serviks, untuk mengurangi jumlah data berlabel dari masing-masing pakar, diusulkan strategi pembelajaran transfer berbasis regulasi yang mendorong model sumber dan target untuk berbagi tanda-tanda koefisien yang sama. Penelitian lain oleh Dewi Sulistyawati, et al [8] dimana pada penelitian tersebut merupakan penelitian observasional analitik dengan desain *case control* dengan kasus pasien dengan diagnosis kanker serviks. Data yang ada di analisis secara deskriptif dan menggunakan analisis *bivariate* dengan hasil 39 responden memiliki hampir persentase umur koitarke yang sama di Poliklinik Onkologi Terpadu (POSA rawat jalan) coiarche awal sebesar 43,6% dan coiarche akhir sebesar 56,4%. Usia koitarke dini di Obstetri dan Pasien rawat jalan ginekologi, sebagai kelompok kontrol, jarang (4 dari 39 responden; 10,3%) dan sebagian besar mereka melakukan debut seksual pada usia >18 tahun (89,7%). Tes chi-kuadrat menunjukkan di sana ada korelasi antara usia koitarke dan kanker serviks ($P=0.002$; $OR=6.76$; $CI\ 95\% \ 2.01-22.75$). Meriam Y Salib, et al [9] dalam penelitiannya menyebutkan kanker serviks adalah kanker paling umum keempat pada wanita dari segala usia di seluruh dunia. Penulis menjelaskan perubahan telah dibuat untuk mencerminkan praktik klinis umum, membedakan hasil prognostik, dan memandu stratifikasi pengobatan. Pilihan pengobatan tergantung pada stadium penyakit dan termasuk pilihan bedah hemat kesuburan dan non-kesuburan serta kemoradioterapi untuk penyakit lanjut lokal. Memahami teknik radiologi serta diagnosis yang digunakan serta jebakan pencitraan umum memastikan staging penyakit yang akurat dan optimalisasi pengobatan.

Kualitas data mining yang diperoleh dari *dataset* yang tidak lengkap dapat dipertanyakan. Selain itu, sangat sulit untuk berhasil memproses dalam *dataset* lengkap.

Dalam proses tersebut terkadang terdapat data yang telah hilang. Beberapa algoritma diusulkan untuk menangani nilai atau data yang telah hilang atau dikenal dengan *missing value*. Output untuk nilai-nilai yang telah hilang yang merupakan hasil dari algoritma imputasi sangat bergantung pada data yang telah lengkap atau bisa disebut data tersebut mengandung data atau informasi *noise*. Kegiatan *preprocessing* sangat berpengaruh terhadap proses klasifikasi yang juga merupakan langkah awal dalam data mining. *Instance selection* merupakan bagian dari *preprocessing* yang memiliki tujuan mengurangi ukuran data dengan menyaring data *noise* dari data yang diberikan [8]. *Instance selection* merupakan algoritma pemilihan yang diterapkan pada sejumlah himpunan bagian terpisah dari *dataset* yang juga merupakan partisi dari data yang tersedia. *Instance selection* sebagai metode untuk pemilihan berdasarkan prototipe dan *instance selection* untuk mendapatkan pelatihan yang ditetapkan untuk algoritma pembelajaran.

Berdasarkan argumentasi tersebut, dapat ditarik kesimpulan bahwa akurasi klasifikasi kanker serviks masih bisa ditingkatkan dengan menggunakan metode *C4.5* yang di melibatkan *instance selection* dengan metode *Naïve Bayes*. *Dataset* yang berukuran besar atau memiliki kelas atau *attribute* yang relatif banyak dan memiliki data yang hilang atau disebut dengan *missing value* sering memiliki *noise* atau *error* pada data [10]. *Instance selection* digunakan pada langkah *pre-processing* data, yang diterapkan untuk mengurangi *dataset* asli ke volume yang dapat dikelola, yang mengarah ke pengurangan sumber daya komputasi yang diperlukan. *Instance selection* disini juga berguna menghapus data *noise* dan *missing value* sebelum melakukan proses klasifikasi, sedangkan pemilihan *naïve bayes* pada *instance selection* digunakan sebagai pemecah masalah untuk pada data yang hilang (*missing value*) dan menimbulkan *noise* pada data karena atribut yang terlalu banyak atau data terlalu besar dan *missing value*, sehingga hasil yang lebih baik bisa didapatkan untuk mencapai tingkat akurasi yang lebih tinggi dari penelitian sebelumnya [11]. Penelitian ini memiliki tujuan untuk melakukan evaluasi terhadap algoritma *Naïve Bayes* sebagai *instance selection* pada klasifikasi kanker serviks.

METODE

Metode yang digunakan pada penelitian ini adalah *Tree C4.5*. Pada penelitian sebelumnya yang dilakukan dengan metode klasifikasi diantaranya *Decision Tree*, *Logistic Regression*, *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN). *Decision Tree* merupakan sebuah diagram alir yang berbentuk seperti struktur pohon yang mana setiap internal *node* menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan output dari pengujian tersebut dan *leaf node* menyatakan kelas-kelas atau distribusi kelas.

Decision Tree mengacu pada struktur hirarkis masalah yang memiliki *node root* & beberapa *node* lainnya. *Node* tingkat menengah disebut *node* pengujian & *node* daun disebut *node* terminal serta *node* keputusan. Kemampuan pohon keputusan adalah dapat mengilustrasikan atau menggambarkan keputusan di antara atribut yang berbeda. Tujuannya adalah untuk memprediksi tahap atribut target tahap kanker serviks berdasarkan beberapa variabel input [12]. *Logistic Function* atau disebut dengan *Logistic Regression* merupakan teknik klasifikasi terkenal yang biasa digunakan dalam ilmu statistik, *machine learning* dan pengetahuan data mining. *Logistic Function* adalah metode klasifikasi statistik terkenal guna pemodelan data biner atau dikotomis. *Logistic Function* memiliki *Value* dan dapat digunakan

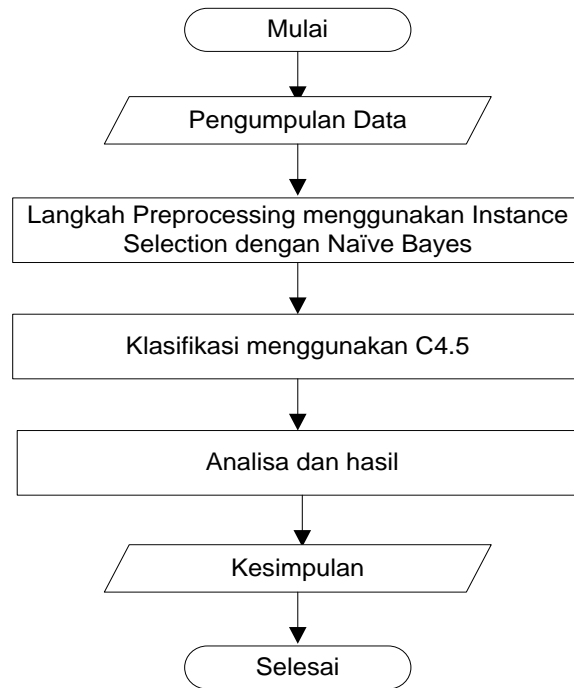
secara luas dalam statistik, *machine learning*, dan komunitas klasifikasi data. *Logistic Function* sebagian besar digunakan dalam masalah klasifikasi biner dalam ilmu terapan kedokteran, biologi dan epidemiologi. Klasifikasi ini merupakan metode klasifikasi statistik standar yang sangat sesuai untuk model yang melibatkan kondisi penyakit (sehat/tidak), pengambilan keputusan (ya/tidak), kematian (mati/hidup).

Klasifikasi *Naïve Bayes* merupakan klasifikasi atau penggolongan probabilistik sederhana yang berdasarkan penerapan teorima bayes dengan asumsi independensi yang kuat. Keuntungan dari klasifikasi *Naïve Bayes* adalah hanya membutuhkan sejumlah kecil data pelatihan untuk memperkirakan nilai *mean* dan varians dari variabel yang diperlukan untuk klasifikasi [13]. *Instance selection* merupakan langkah *preprocessing* penting yang bisa diterapkan dalam *machine learning*. Karena peningkatan ukuran *dataset*, seperti pemilihan yang diterapkan untuk mengurangi data ke volume yang dapat dikelola, yang mengarah kepada pengurangan sumber daya komputasi dalam *machine learning*. *Instance selection* merupakan salah satu cara yang efektif guna mengolah data besar. *Instance selection* berguna untuk mengurangi data atau *scalling down* dengan tetap mempertahankan esensi dari data. Dari perspektif *data mining*, hasil yang ideal untuk *instance selection* adalah poin dari data asli harus menghasilkan hasil atau kesimpulan yang sama dengan seluruh *dataset* yang digunakan[14][15].

Dari beberapa metode yang dilakukan *Decision Tree* merupakan metode terbaik dari penelitian tersebut. Pada penelitian ini penulis menggunakan metode *Tree C4.5*, *Logistic Function* dan *ZeroR* dengan menambahkan *instance selection Naïve Bayes* untuk menciptakan akurasi yang lebih baik dari peneltian sebelumnya. *Instance selection Naïve Bayes* disini digunakan untuk mengurangi *noise* pada *dataset* kanker serviks.

Hasil dari *preprocessing* selanjutnya digunakan untuk diuji menggunakan algoritma C4.5. Proses uji *dataset* kanker serviks tersebut dilakukan dengan menggunakan *tools* Weka. Model yang diusulkan atau diajukan menggambarkan arah serta alur metode yang akan menjelaskan cara kerja atau alur model yang akan diusulkan. Data yang telah diolah akan menjadi model yang sedang diusulkan tersebut. Tahapan penelitian yang dilakukan adalah dengan mengumpulkan *dataset* dari UCI *dataset* yang berupa data kanker serviks. Selanjutnya dilakukan *preprocessing* data dengan *convert* data pada excel, kemudian normalisasi data sampai pada tahap pengklasifikasian menggunakan C4.5 seperti disajikan pada Gambar 1. Pada Gambar 1 dapat diketahui bahwa *dataset* yang digunakan merupakan *dataset* kanker serviks yang dilakukan *preprocessing instance selection* dengan menggunakan *Naïve Bayes*. kemudian dilakukan klasifikasi untuk mengetahui akurasi dengan tahapan sebagai berikut:

1. *Preprocessing* pada penelitian ini adalah melibatkan *instance selection* untuk mengurangi *missing value* pada *dataset* yang ada.
2. Data yang sudah melewati proses *preprocessing* akan dilakukan proses klasifikasi menggunakan tiga metode yang nantinya akan dibandingkan dengan mencari nilai akurasi paling maksimal.
3. Penerapan metode klasifikasi menggunakan metode C4.5, *Logistic Regression* atau *Logistic Function* dan *ZeroR*
4. Evaluasi yang digunakan adalah menggunakan *accuracy*.



Gambar 1. Kerangka Pemikiran Penelitian

Data yang digunakan adalah data publik, “*Cervical Cancer (Risk Factors) Dataset*” yang diambil dari website *Machine Learning Repository* (UCI) [16] yang berisi daftar risiko kanker serviks yang berfokus pada prediksi indikator atau diagnosis kanker serviks dengan fitur-fitur yang tersedia. Adapun fitur-fitur itu mencakup informasi demografis, kebiasaan dan sejarah *medical record*. Data tersebut didapatkan dari Hospital Universitario di Caracas, Venezuela.

Dataset tersebut terdiri dari 858 data pasien dengan melibatkan 32 atribut dan 4 atribut yang diantaranya terdiri dari *Age*, *Number of secyal intercourse*, *First sexual intercourse*, *Num of Pregnancies*, *Smoke*, *Smoke(years)*, *Smokes (Pack/Year)*, *Hormonal Contraceptives*, *Hormonal Contraceptives (Years)*, *IUD*, *IUD (Years)*, *STDs*, *STDs (number)*, *STDs:condylomatosis*, *STDs: cervical condylomatosis*, *STDs: Vaginal condylomatosis*, *STDs:vulvo-perineal condylomatosis*, *STDs:syphilis*, *STDs:pelvic inflammatory disease*, *STDs:genital herpes*, *STDs Molluscum contagiosum*, *STDs: AIDS*, *STDs:HIV*, *STDs:Hepatitis B*, *STDs:HPV*, *STDs:Number of Diagnosis*, *STDs:Time since first diagnosis*, *STDs: Time since last diagnosis*, *Dx:Cancer*, *Dx:CIN*, *Dx: HPV*, *DX*, *Hinselmann*, *Schiller*, *cytology* dan *Biopsy*[15].

Karakteristik *dataset* yang digunakan disajikan pada Tabel 1 dengan tipe *multivariate* yang memiliki atribut *integer* dan *real* yang diklasifikasikan pada penyakit kanker serviks dan memiliki *missing value*. Pada penelitian ini terdapat empat *class* yaitu *Hinselmann*, *Schiller*, *Cytology* dan *Biopsy*, namun pada penelitian ini akan terfokus dan terbatas pada label atau kelas biopsi saja.

Tabel 1. Dataset Kanker Serviks

<u>Attribute</u>	<u>Type</u>	<u>Attribute</u>	<u>Type</u>
- Age	Integer	- STDs: syphilis	Boolean
- Number of sexual partener	Integer	- STDs: pelvic inflammatory disease	Boolean
- First sexual intercourse (age)	Integer	- STDs: genital herpes	Boolean
- Number of peregnancies	Integer	- STDs: Molluscum contagiosum	Boolean
- Smokes	Boolean	- STDs: AIDS	Boolean
- Smokes (years)	Boolean	- STDs: HIV	Boolean
- Smokes (pack/years)	Boolean	- STDs: Hepatitis B	Boolean
- Hormonal contraceptives	Boolean	- STDs: HPV	Boolean
- Hormonal contraceptives (years)	Integer	- STDs: Number of diagnosis	Integer
- IUD	Boolean	- STDs:Time since first diagnosis	Integer
- IUD (years)	Boolean	- STDs:Time since last diagnosis	Integer
- STDs	Integer	- Dx:cancer	Boolean
- STDs (number)	Boolean	- Dx:CIN	Boolean
- STDs:condylomatosis	Integer	- Dx: HPV	Boolean
- STDs: cervical condylomatosis	Boolean	- Dx	Boolean
- STDs: vaginal condylomatosis	Boolean		Boolean
- STDs: vulvo-perineal condylomatosis	Boolean		Boolean

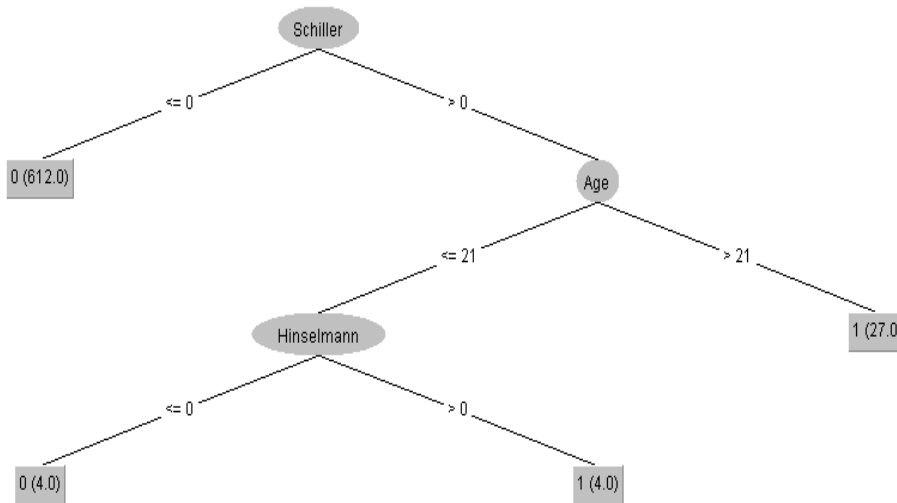
Tabel 2 menjabarkan bahwa *dataset* memiliki karakteristik atribut berupa integer dan real dengan data sebanyak 858 data pasien dan jumlah atribut sebanyak 36 dan terdapat *missing value*.

Tabel 2. Deskripsi Dataset

Dataset Characteristics	Multivariate	Number of Instances	858
Attribute Characteristics	Integer, Real	Number Of Atributes	36
Associated Tasks	Classification	Missing Values	Yes

HASIL DAN PEMBAHASAN

Dalam pengujian kanker serviks menggunakan dataset yang berasal dari UCI *Machine Learning Repository* yang dilakukan dengan metode klasifikasi C4.5 dengan beberapa eksperimen yang memiliki tujuan yaitu untuk mendapatkan nilai akurasi maksimal menggunakan *tools* Weka memiliki hasil algoritma C4.5 mendapatkan hasil akurasi relative tinggi yaitu 99,69% dengan nilai *Kappa Statistic* sebesar 0,9671, *Total Number of Instance* 647 dan *Incorrectly Classified Instance* dan *Incorrectly Classified Instance* sebesar masing-masing 2 dan 645. menghasilkan *recall* pada *class* 1 yaitu 1,000 atau 100% sedangkan *recall* dengan *class* 0 adalah 0,997 atau 99,7%. Klasifikasi ini menghasilkan pohon keputusan seperti Gambar 2.



Gambar 2. Pohon Keputusan C4.5

Berdasarkan hasil pohon keputusan pada Gambar 2. diperoleh bahwa dari beberapa keterangan atau deskripsi atribut beserta isi yang ada pada data, yang harus melakukan skrining kanker serviks sebanyak 27 pasien yang semuanya memiliki usia di atas 21 tahun. Hasil pengujian kedua dengan dataset yang sama yaitu kanker serviks dataset dengan menerapkan metode klasifikasi *Logistic Function* memiliki nilai akurasi sebesar 99,38%. Dengan menerapkan metode klasifikasi *ZeroR* diperoleh nilai akurasi sebesar 95,20% dengan nilai *Kappa Statistic* dan *Time to Build Model* sebesar 0, *Total Number of Instance* sebesar 647, *Correctly Classified Instances* 616 dan *Incorrectly Classified Instances* sebesar 31. Perbandingan hasil nilai akurasi dengan 3 metode disajikan pada Tabel 3.

Tabel 3. Nilai Akurasi Masing-masing Klasifikasi

	<i>C4.5</i>	<i>Logistic Function</i>	<i>ZeroR</i>
Accuracy	99,69%	99,38%	95,20%

Berdasarkan Tabel 3 dapat dinyatakan bahwa dataset kanker serviks dari UCI *Machine Learning Repository* dengan melibatkan 858 pasien dan 36 atribut dan prediksi klasifikasi *C4.5*, *Logistic Function* dan *ZeroR* didapatkan hasil bahwa klasifikasi terbaik adalah dengan metode klasifikasi *C4.5* dengan nilai akurasi sebanyak 99,69%. Kanker serviks merupakan penyebab kematian yang signifikan. Seperti penyebab pada beberapa penyakit lain, keberadaan beberapa metode skrining atau diagnosis serta referensi beberapa dokter yang subjektif menciptakan ekosistem yang kompleks untuk metode otomatisasi. Dengan hasil yang diberikan dan akurasi yang ada yaitu dari *instance selection* dengan *Naïve Bayes* menggunakan klasifikasi *C4.5* cocok digunakan dalam kasus ini. Pohon keputusan yang ada memberikan hasil dari beberapa ciri-ciri yang di alami oleh pasien seperti *Age*, *Number of secyal intercourse*, *First sexual intercourse*, dan lain-lain.

KESIMPULAN

Pada penelitian ini dilakukan dengan menggunakan langkah *preprocessing* yang melibatkan *instance selection* dengan *Naïve Bayes* guna mendapatkan akurasi terbaik dengan menyeleksi data yang masih *missing value*. Klasifikasi terbaik didapatkan dari klasifikasi C4.5 yaitu sebesar 99,69% dengan *Kappa Statistic* sebesar 0,9671 *Correctly Classified Instances* senilai 645 dan *Incorectly Classified Instance* sebesar 2. Dari hasil tersebut dapat disimpulkan bahwa algoritma C4.5 merupakan algoritma terbaik untuk menciptakan akurasi yang maksimal dibandingkan dengan algoritma *Logistic Function* dan *ZeroR*. Pada penelitian selanjutnya di harapkan mampu mendapatkan hasil yang lebih baik serta bisa menambahkan beberapa label lain dengan data yang sama mengingat penelitian ini hanya fokus atau memiliki batasan masalah pada satu label saja yaitu *biopsy*.

DAFTAR PUSTAKA

- [1] S. Athinarayanan, Srinath M.V., “Classification Of Cervical Cancer Cells In Pap Smear Screening Test”, ICTACT Journal on Image and Video Processing, vol. 06, 2016.
- [2] N. Kumar, “Cervical Cancer; a Nightmare for Womanhood: Review of Recent Advances”, Journal Women’s Health and Gynecology, 2016.
- [3] P. Tsikouras, S. Zervoudis, B. Manav, E. Tomara, G. Iatrakis, C. Romanidis, A. Bothou, G. Galazios, “Cervical Cancer: Screening, Diagnosis and Staging”, Journal of The Balkan Union of Oncology, 2016.
- [4] D. Sofjayana Latha *et al.*, “Staging Prediction in Cervical Cancer Patients- A Machine Learning Approach”, International Journal of Innovative Research and Practices, vol. 2, 2014.
- [5] Sunny Sharma, “Cervical Cancer Stage Prediction Using Decision Tree Approach of Machine Learning”, International Journal of Advanced Research in Computer and Communication Engineering, vol. 5, 2016.
- [6] Y. M. S. Al-Wesabi, A. Choudhury, “Classification Of Cervical Cancer Dataset”, IISE Annual Conference, 2018.
- [7] K. Fernandes, J. S. Cardoso, J. Fernandes, Transfer Learning with Partial Observability Applied to Cervical Cancer Screening, Iberian Conference on Pattern Recognition and Image Analysis. Springer International Publishing, 2017
- [8] D. Sulistyawati, Z. Faizah, E.M. Kurniawati, “An Association Study of Cervical Cancer Correlated With The Age of Coitarche in Dr. Soetomo Hospital Surabaya”, Indonesian Journal of Cancer, 2020.
- [9] M. Y. Salib, J. H. B. Russel, V. R. Stewart, S.A. Sudderuddin, D. Barwick, A.G. Rockall, N. Bharwati, “2018 FIGO Staging Classification for Cervical Cancer: Added Benefits of Imaging”, RSNA, 2018
- [10] Y. Song, J. Liang, J. Lu, X. Zhao, “An Efficient Instance selection Algorithm for kNearest Neighbor Regression”, Neurocomputing, Vol. 251, No. 26-34, 2017
- [11] N. Jankowski, M. Grochowski, “Comparison of Instances Selection Algorithms Survey”, International Conference on Artificial Intelligence and Soft Computing, 2004.
- [12] J. Han and M. Kamber, “Data Mining Concepts and Techniques”, Morgan Kaufman Publishers, USA pp 279-322, 2003.
- [13] Z. K. Sentruk, R. Kara, “Breast Cancer Diagnosis Via Data Mining: Performance Analysis of Seven Different Algorithm”, Computer Science & Engineering: An International Journal, Vol. 4, No. 1, 2014
- [14] C.G. Osorio, A.D. H. Garcia, N.G. Pedrajas, “Democratic Instance Selection: A Linier Complexity Instance selection Algorithm Based On Classifier Ensemble Concept”,

- Artificial Intelligent, Vol. 174, No.410, 2010
- [15] N. G.Pedrajas, A. D.H. Garcia, “Boosting Instance selection Algorithms”, Journal Knowledge based System, 2014
- [16] <https://archive.ics.uci.edu/ml/datasets/Cervical+cancer+%28Risk+Factors%29> , diakses pada 10 September 2021



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)
