

Evaluasi Identifikasi Kanker Serviks Berdasarkan Data Risiko Perilaku dengan Data Mining

✉¹Edi Jaya Kusuma, ²Ririn Nurmandhani, ²Sri Handayani

¹Program Studi Rekam Medis dan Informasi Kesehatan, Fakultas Kesehatan, Universitas Dian Nuswantoro Semarang, 50131, Indonesia

²Program Studi Kesehatan Masyarakat, Fakultas Kesehatan, Universitas Dian Nuswantoro Semarang, 50131, Indonesia

ABSTRAK

Kanker Serviks merupakan salah satu jenis kanker yang disebabkan oleh Human Papillomavirus atau HPV. Di Indonesia, kasus kemunculan kanker serviks berada di peringkat ke-2 dibawah kanker payudara. Kebutuhan akan deteksi dini kanker serviks sangat diperlukan, terlebih kemunculan kanker serviks dapat dikenali ketika kondisi kanker memasuki stadium akhir. Dengan memanfaatkan teknologi serta berdasarkan data perilaku, penelitian ini mengusulkan identifikasi dini kanker serviks menggunakan kombinasi seleksi fitur information gain dan data mining. Implementasi dilakukan pada dataset Cervical Cancer Risk Behavioral. Metode information gain mampu menghasilkan 9 fitur utama yang akan digunakan dalam tahap evaluasi dengan data mining. Dari hasil evaluasi beberapa model data mining diketahui metode Naive Bayes mampu memberikan performa terbaik dengan pencapaian 93,21% akurasi, 96% sensitivitas, dan 83,33% spesifisitas. Sehingga dapat disimpulkan bahwa penggunaan skema 9 atribut berdasarkan bobot information gain mampu memberikan peningkatan kemampuan model data mining dalam mengidentifikasi kemunculan kanker serviks.

Kata Kunci: Data Mining, Kanker Serviks, Perilaku, Seleksi Fitur.

Evaluation of Cervical Cancer Identification based on Behavioral Risk Dataset using Data Mining

ABSTRACT

Cervical cancer is a type of cancer caused by the Human Papillomavirus or HPV. In Indonesia, the incidence of cervical cancer is ranked 2nd under the breast cancer. The need for early detection of cervical cancer is necessary because the emergence of cervical cancer can only be identified when the cancer is in a late stage. By utilizing technology and based on behavioral data, this study proposes early detection of cervical cancer using a combination of information gain feature selection and data mining. The implementation is carried out on the Cervical Cancer Risk Behavioral dataset. The information gain method provides 9 main features which will be used in the evaluation phase. From the evaluation results, it can be seen that the Naive Bayes method is able to provide the best performance with the value of 93.21% accuracy, 96% sensitivity, and 83.33% specificity. Therefore, it can be concluded that the 9 attributes scheme is able to increase the capability of data mining models in identifying the cervical cancer.

Keywords: Behavior, Cervical Cancer, Data Mining, Feature Selection.

PENDAHULUAN

Kanker Serviks merupakan salah satu jenis kanker yang disebabkan oleh Human Papillomavirus atau HPV (Lu dkk., 2021). Pada dasarnya kemunculan virus ini ditandai dengan adanya kutil di permukaan kulit di area kelamin, namun ada pula proses infeksi HPV yang tidak menimbulkan gejala. Proses perpindahan virus umumnya dapat terjadi pada saat hubungan seksual. Berdasarkan data yang dihimpun oleh GLOBOCAN (Global Cancer Statistic), tercatat pada tahun 2020 kasus kanker serviks pada wanita mencapai 604.127 kasus dengan kasus kematian hingga lebih dari 341.831 kasus (Sung dkk., 2021). Selain itu, di Indonesia, kasus kemunculan kanker serviks berada di peringkat ke-2 dibawah kanker payudara, dimana angka kejadian kanker serviks sebesar 23,4 per 100.000 penduduk dengan rata-rata kasus mortalitas sebesar 90-100 per 100.000 penduduk (Riani dan Ambarwati, 2020).

Tingginya kasus mortalitas pada pasien yang mengidap kanker serviks disebabkan oleh kesadaran pasien terhadap kemunculan kanker serviks yang baru terlihat saat memasuki stadium lanjut atau akhir. Salah satu upaya dalam mereduksi jumlah pasien kanker serviks adalah dengan melakukan deteksi dini terhadap kemunculan kanker serviks. Beberapa penelitian mengusulkan metode-metode ataupun pendekatan untuk mendeteksi kemunculan kanker serviks sejak dini. Salah satunya adalah penelitian yang dilakukan oleh Riani dan Ambarwati (Riani dan Ambarwati, 2020) dimana pada penelitian tersebut dilakukan pemaparan dan diskusi materi mengenai deteksi dini kanker serviks serta demonstrasi mengenai pemeriksaan IVA (inspeksi visual asam asetat) dan *pap smear* terhadap masyarakat di wilayah kabupaten Banyumas, Jawa Tengah. Dari hasil evaluasi kegiatan tersebut didapatkan bahwa terdapat peningkatan pengetahuan mengenai gejala kanker serviks serta peningkatan kesadaran masyarakat terhadap pemeriksaan IVA dan *pap smear*. Kemudian dengan memanfaatkan perkembangan teknologi Agustyawati dkk. (2021), mengusulkan rancangan aplikasi deteksi dini kanker serviks memanfaatkan teknik *deep learning Convolutional Neural Network* (CNN). Pada penelitian tersebut

memanfaatkan data citra digital berupa hasil IVA yang dibagi menjadi dua kelas yaitu serviks positif sebesar 77 citra dan serviks negatif sejumlah 82 citra. Dari hasil evaluasi didapatkan bahwa rancangan aplikasi tersebut mampu mendeteksi kanker serviks dengan akurasi sebesar 96%.

Selain menggunakan pemeriksaan fisik, deteksi dini kanker serviks dapat diidentifikasi dengan melihat faktor eksternal seperti perilaku (*behavior*) maupun lingkungan (*environment*) dari seseorang. Salah satunya yang dilakukan oleh Shalikhah dkk. (2021), yang melakukan pengujian korelasi antara dukungan keluarga (*environmental support*) dengan sikap deteksi dini kanker serviks. kemudian menurut hasil analisa dari penelitian Fransisca dan Adhistry (2021), dukungan keluarga merupakan faktor yang sangat penting dalam menangani dan meningkatkan motivasi pasien penderita kanker serviks. Selanjutnya penggunaan faktor perilaku juga dilakukan oleh Alpen (2021), dimana pada penelitian tersebut memanfaatkan data perilaku sebagai dasar dalam menentukan kemunculan kanker serviks pada pasien. Penelitian tersebut melakukan uji coba terhadap 8 (delapan) algoritma klasifikasi untuk mengidentifikasi kemunculan kanker serviks berdasarkan data perilaku pasien. Dengan memanfaatkan *tools* WEKA, penelitian tersebut menyimpulkan bahwa metode *Support Vector Machine* (SVM) mampu memperoleh hasil akurasi sebesar 91,67%.

Berdasarkan latar belakang diatas, penulis mengusulkan evaluasi terhadap deteksi dini kanker serviks berdasarkan data resiko perilaku dengan memanfaatkan teknik data mining. Evaluasi dilakukan dengan menganalisa faktor-faktor perilaku untuk mengetahui tingkat relevansinya terhadap kemunculan kanker serviks. Pada penelitian ini, proses analisa dilakukan dengan memanfaatkan teknik pembobotan *information gain* (IG). Kemudian dari hasil analisa tersebut, dilakukan prediksi menggunakan beberapa teknik *machine learning* untuk mengetahui performa dari faktor-faktor yang terpilih. Pada penelitian ini, teknik *machine learning* yang digunakan adalah *K-Nearest Neighbor* (KNN), *Naïve Bayes* (NB), *Decision Tree* (DT), *Logistic Regression*, dan *Support Vector Machine*

(SVM). Dari hasil pengujian dan evaluasi tersebut diharapkan dapat menjadi referensi dalam pengembangan suatu *machine learning* yang mampu memberikan rekomendasi terhadap kemunculan kanker serviks pada seseorang berdasarkan pola kebiasaan atau perilakunya.

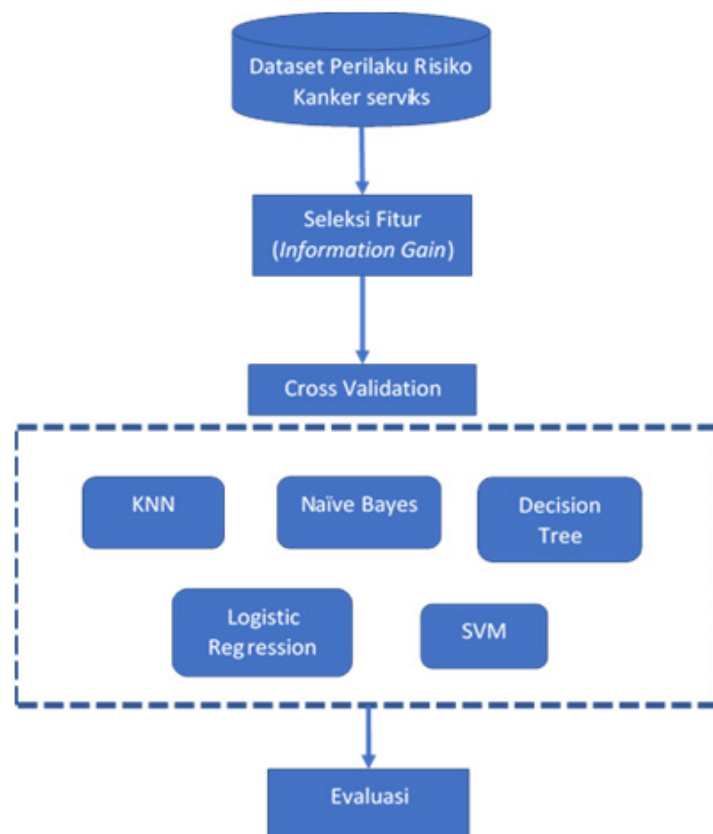
METODE PENELITIAN

Pada penelitian ini digunakan metode penelitian deskriptif kuantitatif dimana penulis melakukan analisa dengan menggunakan pendekatan statistic dan data mining yang kemudian hasilnya akan dideskripsikan. Seperti pada Gambar 1, penelitian ini memanfaatkan dataset sekunder yang berjudul *Cervical Cancer Behavioral Risk Dataset* serviks (Sobar dkk., 2016). Dataset tersebut disediakan secara publik (*open access*) melalui *web* dapat diakses melalui <https://archive.ics.uci.edu/ml/datasets/Cervical+Cancer+Behavior+Risk>. Dataset ini memiliki 72 data dengan 18 atribut (kolom) dan

sebuah atribut yang menjadi label yang berisi angka 0 (nol) sebagai pasien Bukan Kanker Serviks dan 1 (satu) yang merepresentasikan pasien Kanker Serviks.

Dari dataset tersebut, kemudian dilakukan analisa dengan menggunakan teknik seleksi fitur *information gain* untuk mencari atribut yang memiliki relevansi terhadap kanker serviks. *Information gain* memanfaatkan teknik *scoring* dalam membobotkan suatu atribut dengan memilih nilai *entropy* maksimal (Alelyani, 2021).

Dari hasil seleksi fitur tersebut, kemudian dilakukan implementasi beberapa metode data mining yaitu KNN, *Naïve Bayes*, *Decision Tree*, *Logistic Regression*, dan SVM dengan memanfaatkan metode *Cross Validation*. Setelah itu hasil evaluasi dideskripsikan untuk mengetahui pola dan performa akurasi (*aaccuracy*), sensitivitas (*sensitivity*), dan spesifisitas (*specificity*) dari proses identifikasi



Sumber: Data Primer Diolah, 2022

Gambar 1
Bagan Alur Metode Penelitian yang Diusulkan

kanker serviks. Nilai akurasi (*accuracy*) menunjukkan kemampuan suatu model data mining dalam memprediksi kategori atau label dari suatu data secara benar. Semakin tinggi semakin akurat suatu model dalam prediksi kategori atau label dari data yang diujikan.

Kemudian, nilai sensitivitas (*sensitivity*) menunjukkan kemampuan suatu model dalam menunjukkan suatu data berlabel negatif diantara data yang benar-benar berlabel negatif. Sebaliknya, nilai spesifisitas (*specificity*) menunjukkan kemampuan suatu model dalam menentukan suatu data berlabel positif diantara data-data yang benar-benar berlabel positif (Sarosa dkk., 2019).

HASIL DAN PEMBAHASAN

Evaluasi identifikasi kemunculan kanker serviks berdasarkan dataset perilaku risiko kanker serviks telah dilakukan. Pada dataset perilaku risiko kanker serviks (*Cervical Cancer Behavioral Risk Dataset*) memiliki 18 atribut utama.

Proses evaluasi diawali dengan melakukan pembobotan pada masing-masing atribut pada Tabel 1 dengan menggunakan teknik *information gain*. Proses pembobotan termasuk dalam proses evaluasi multivariat dimana masing-masing atribut akan diuji relevansinya dengan kemunculan kanker serviks. Pada teknik *information gain* nilai distribusi bobot merupakan nilai decimal yang berada diantara 0 dan 1 dengan kriteria semakin besar nilai bobot suatu atribut, maka semakin relevan atribut tersebut terhadap atribut *dependent*-nya.

Pada Tabel 2, representasi atribut diurutkan dari bobot terbesar hingga terkecil. Dari hasil tersebut dapat dilihat bahwa dari 18 atribut yang dievaluasi menggunakan teknik *information gain* terdapat tiga atribut dengan bobot tertinggi yaitu atribut Keparahan (*severity*) dengan nilai 0,2739 kemudian diikuti oleh atribut kemampuan (*abilities*) dengan bobot 0,2438, dan kekuatan (*strength*) dengan bobot 0,2236. Dari hasil pembobotan diatas,

Tabel 1
Daftar Atribut dari *Cervical Cancer Behavioral Risk Dataset*

No.	Atribut	Variabel
1	Makan (<i>eating</i>)	Perilaku (<i>behavior</i>)
2	Kebersihan Pribadi (<i>personal hygiene</i>)	
3	Agregasi (<i>aggregation</i>)	
4	Komitmen (<i>commitment</i>)	Niat (<i>intention</i>)
5	Konsistensi (<i>consistency</i>)	
6	Spontanitas (<i>spontaneity</i>)	Sikap (<i>attitude</i>)
7	Orang Terdekat (<i>significant person</i>)	
8	Pemenuhan (<i>fulfillment</i>)	Norma (<i>norm</i>)
9	Kerentanan (<i>vulnerability</i>)	
10	Keparahan (<i>severity</i>)	Persepsi (<i>perception</i>)
11	Kekuatan (<i>strength</i>)	
12	Kemauan (<i>willingness</i>)	Motivasi (<i>motivation</i>)
13	Emosional (<i>emotional</i>)	
14	Apresiasi (<i>appreciation</i>)	Dukungan Sosial (<i>social support</i>)
15	Instrumental (<i>instrumental</i>)	
16	Pengetahuan (<i>knowledge</i>)	Pemberdayaan (<i>empowerment</i>)
17	Kemampuan (<i>abilities</i>)	
18	Keinginan (<i>desires</i>)	

Sumber: <https://archive.ics.uci.edu/ml/datasets/Cervical+Cancer+Behavior+Risk>

besarnya bobot pada masing-masing atribut menunjukkan tingkat relevansi atribut terhadap label atau atribut *dependent*-nya, sehingga dapat dikatakan bahwa atribut keparahan, kemampuan, dan kekuatan merupakan tiga atribut yang memiliki relevansi yang tinggi terhadap kemunculan kanker serviks. Di lain sisi, atribut spontanitas (*spontaneity*), konsistensi (*consistency*), dan instrumental (*instrumental*) merupakan tiga atribut dengan bobot terkecil. Hasil tersebut secara implisit menerangkan bahwa ketiga atribut tersebut memiliki relevansi yang rendah terhadap proses identifikasi kanker serviks pada pasien.

Dari hasil pembobotan menggunakan teknik *information gain*, kemudian dilakukan evaluasi lanjutan menggunakan beberapa teknik klasifikasi yaitu KNN, Naïve Bayes, Decision Tree, Logistic Regression, dan SVM. Proses klasifikasi dilakukan dalam 2 (dua) skema yaitu menggunakan 100% atau keseluruhan atribut (18 atribut) pada dataset *Cervical Risk Behavior* dan 50% atau sebagian dari atribut yang ada pada dataset tersebut. Sebagian atribut (50%) atau 9 atribut yang diambil berdasarkan

hasil pembobotan *information gain*. Selain itu, proses evaluasi juga memanfaatkan metode *cross validation* untuk memastikan bahwa model *data mining* yang nantinya dihasilkan mampu menangani data-data yang belum diketahui (*unknown data*) (Bates dkk., 2021).

Tahap pengujian dilakukan dengan memanfaatkan tools *Rapid Miner* yang merupakan salah satu *software data mining* untuk mengolah data. Berikut merupakan proses evaluasi yang dibangun dan dilakukan dengan menggunakan *Rapid Miner*.

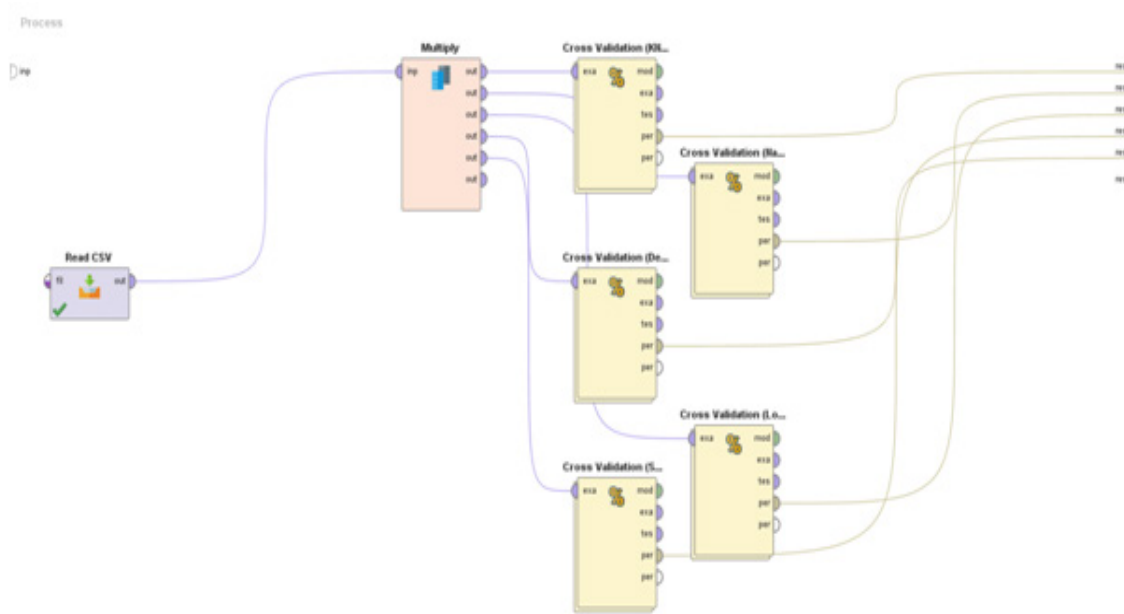
Pada Gambar 2 dapat dilihat sebuah jendela proses dari software *Rapid Miner* yang dimana digunakan untuk mengeksekusi keseluruhan atribut (18 atribut) dari dataset *Cervical Risk Behavior*. Dari gambar tersebut juga terdapat 5 (lima) operator *cross validation* yang masing-masing merepresentasikan model klasifikasi KNN, *Naïve Bayes*, *Decision Tree*, *Logistic Regression*, dan *Support Vector Machine* (SVM). Dari hasil proses yang dijalankan *Rapid Miner*, didapatkan hasil sebagai berikut.

Tabel 2

Daftar Bobot *Information Gain* pada Atribut dari *Cervical Cancer Behavioral Risk Dataset*

Atribut	Bobot IG
Keparahan (<i>severity</i>)	0,2739
Kemampuan (<i>abilities</i>)	0,2438
Kekuatan (<i>strength</i>)	0,2236
Kerentanan (<i>vulnerability</i>)	0,2234
Kemauan (<i>willingness</i>)	0,2223
Pemenuhan (<i>fulfillment</i>)	0,2131
Pengetahuan (<i>knowledge</i>)	0,1907
Keinginan (<i>desires</i>)	0,1907
Emosional (<i>emotional</i>)	0,1895
Kebersihan Pribadi (<i>personal hygiene</i>)	0,1326
Agregasi (<i>aggregation</i>)	0,1033
Apresiasi (<i>appreciation</i>)	0,0847
Makan (<i>eating</i>)	0,0756
Komitmen (<i>commitment</i>)	0,0680
Orang Terdekat (<i>significant person</i>)	0,0677
Instrumental (<i>instrumental</i>)	0,0593
Konsistensi (<i>consistency</i>)	0,0343
Spontanitas (<i>spontaneity</i>)	0,0191

Sumber: Data Primer Diolah, 2022



Sumber: Data Primer Diolah, 2022

Gambar 2
Proses *Rapid Miner* dengan menggunakan Keseluruhan Atribut Dataset (18 Atribut)

Tabel 3
Hasil Evaluasi terhadap Keseluruhan Atribut (18 Atribut)
pada Dataset Menggunakan Metode *Data Mining*

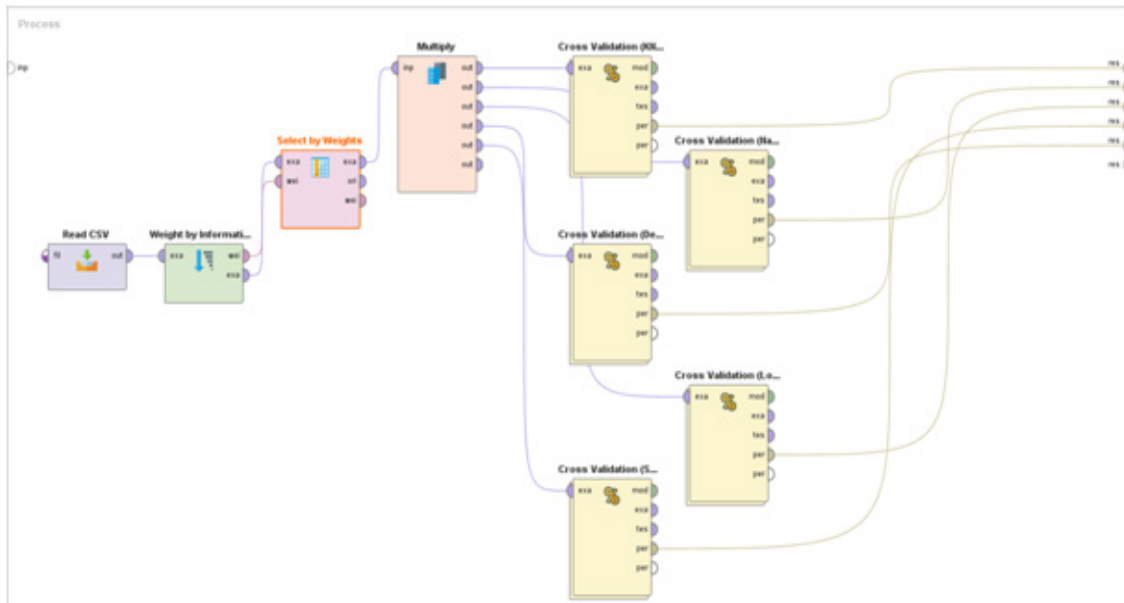
Metode	Akurasi	Sensitivitas	Spesifisitas
KNN	88,93	96,00	71,67
Naïve Bayes	92,14	96,00	85,00
Decision Tree	87,32	92,00	75,00
Logistic Regression	87,50	92,33	73,33
SVM	87,50	94,00	68,33

Sumber: Data Primer Diolah, 2022

Pada Tabel 3, dapat diperhatikan bahwa dari kelima metode yang diujikan, metode *Naïve Bayes* yang mampu menghasilkan skor akurasi, sensitivitas, dan spesifisitas yang tinggi dibandingkan dengan metode lainnya. Pada hasil tersebut dapat diketahui bahwa metode *Naïve Bayes* mampu mengidentifikasi data dengan kelas “Kanker Serviks”, dan “Bukan Kanker Serviks” dengan akurasi sebesar 92,14%. Kemudian metode tersebut juga mampu mengidentifikasi data dengan label “Bukan Kanker Serviks” diantara data-data yang benar-benar berlabel “Bukan Kanker Serviks” yang ditunjukkan dengan nilai sensitivitas sebesar 96%. Lalu berdasarkan nilai spesifisitas, metode *Naïve Bayes* mampu

memprediksi data dengan label “Kanker Serviks” diantara data aktual berlabel “Kanker Serviks” dengan ketepatan sebesar 85%.

Evaluasi pada skema atribut 50% yang diambil dari hasil pembobotan *information gain* menghasilkan atribut yang terdiri dari: Keparahan (*severity*), Kemampuan (*abilities*), Kekuatan (*strength*), Kerentanan (*vulnerability*), Kemauan (*willingness*), Pemenuhan (*fulfillment*), Pengetahuan (*knowledge*), Keinginan (*desires*), dan Emosional (*emotional*). Dengan menggunakan 9 (sembilan) atribut tersebut, maka dapat dihasilkan jendela proses *Rapid Miner* seperti yang ditunjukkan pada Gambar 3.



Sumber: Data Primer Diolah, 2022

Gambar 3
Proses Rapid Miner dengan Menggunakan 50%
Atribut dari Keseluruhan Dataset (9 Atribut)

Tabel 4
Hasil Evaluasi terhadap 50% Atribut (9 Atribut)
pada Dataset menggunakan Metode Data Mining

Metode	Akurasi	Sensitivitas	Spesifisitas
KNN	90,54	96,00	80,00
Naïve Bayes	93,21	96,00	83,33
Decision Tree	88,75	94,00	75,00
Logistic Regression	90,18	92,33	85,00
SVM	91,61	94,00	86,67

Sumber: Data Primer Diolah, 2022

Pada proses *Rapid Miner* diatas, terdapat penambahan operator *weight by information gain* dan *select by weight*. Kedua operator tersebut berguna untuk membobotkan atribut pada dataset dengan teknik *information gain* dan memilih atribut-atribut berdasarkan kriteria yang sudah ditentukan (50% dari keseluruhan atribut dataset). Mirip dengan yang ada pada Gambar 2 (skema 100% atribut), pada skema 50% ini, proses evaluasi dilakukan terhadap 5 (lima) model klasifikasi dengan memanfaatkan teknik *cross validation*.

Dari proses *Rapid Miner* pada Gambar 3 dapat diperoleh hasil evaluasi pada masing-masing model klasifikasi sebagai berikut (Tabel 4).

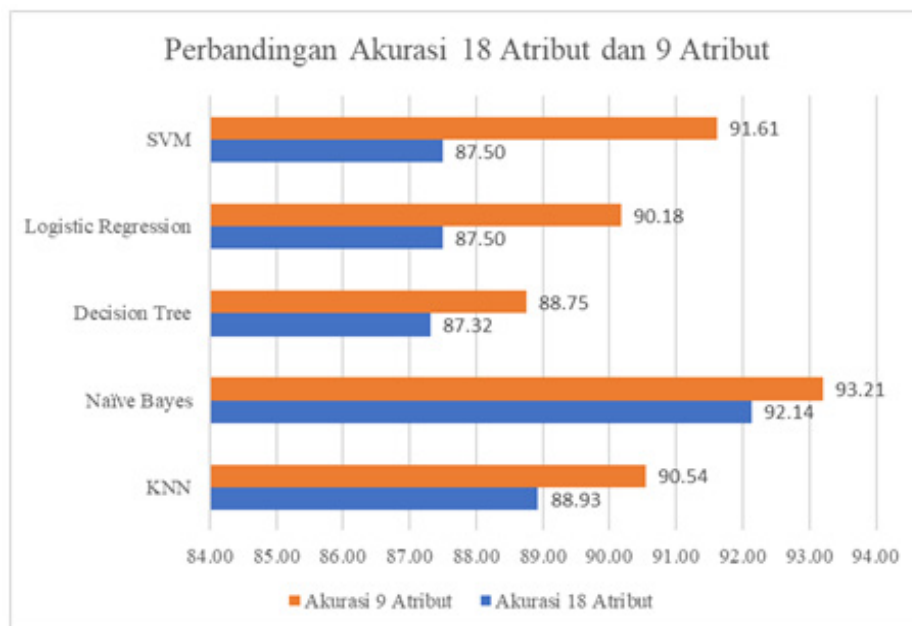
Hasil evaluasi dengan skema 50% atribut dari keseluruhan atribut dataset dapat dilihat pada Tabel 4. Pada tabel tersebut, metode *Naïve Bayes* kembali memperoleh nilai tertinggi untuk akurasi dan sensitivitas dibandingkan dengan metode lain. Metode *Naïve Bayes* mampu memperoleh nilai akurasi sebesar 93,21% disusul oleh metode SVM dengan nilai akurasi sebesar 91,61%. Kemudian untuk nilai sensitivitas, metode *Naïve Bayes* mampu menghasilkan nilai sebesar 96% yang dimana metode KNN juga memperoleh nilai sensitivitas yang sama. Namun untuk nilai spesifisitas, nilai terbaik diraih oleh metode SVM dengan nilai 86,67% diikuti oleh metode *Logistic Regression* dengan nilai sebesar 85%.

Berdasarkan hasil evaluasi dari kedua skema terlihat bahwa terdapat beberapa peningkatan misalnya pada Gambar 4 yang menunjukkan perbandingan nilai akurasi pada masing-masing metode dari kedua skema yang digunakan.

Pada Gambar 4, terlihat bahwa hasil akurasi pada masing-masing metode mengalami kenaikan ketika menggunakan skema 9 atribut. Kenaikan terbesar dicapai oleh metode SVM yang dimana pada skema 18 atribut hanya mendapat akurasi sebesar 87,50%, tetapi meningkat menjadi 91,61% ketika menggunakan skema 9 atribut. Selain itu, akurasi tertinggi dicapai oleh metode *Naïve Bayes*, dimana metode tersebut mampu mencapai nilai akurasi sebesar 93,21% pada skema 9 atribut. Hal ini menunjukkan bahwa pada skema 9 atribut, metode-metode klasifikasi yang diujikan mampu

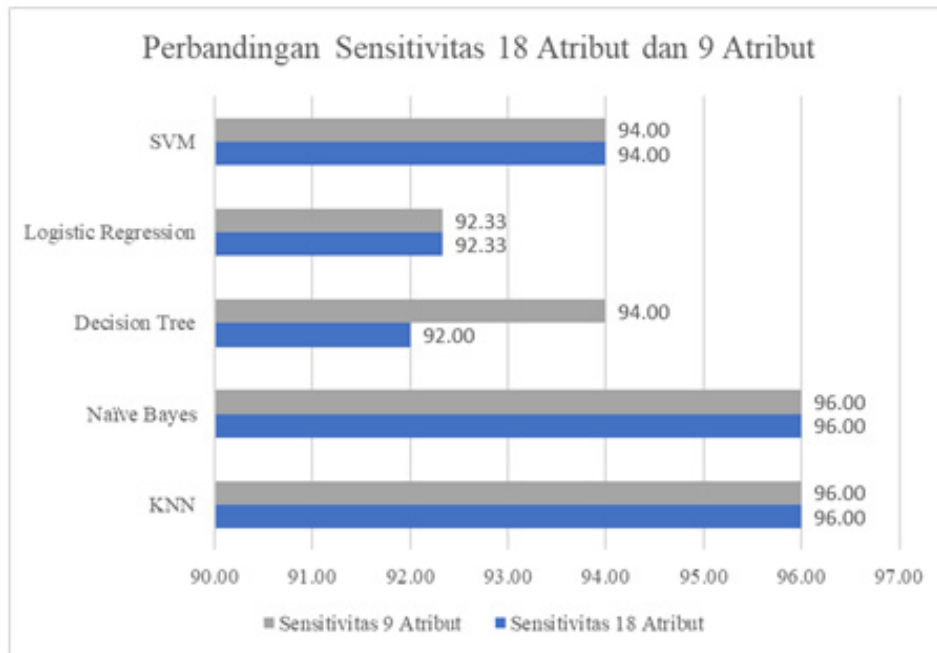
memprediksi data berlabel “Kanker Serviks” dan “Bukan Kanker Serviks” dengan lebih baik dibandingkan dengan skema 18 atribut. Perbedaan hasil dari kedua skema juga terjadi pada nilai evaluasi sensitivitas (Gambar 5) dan spesifisitas (Gambar 6).

Pada evaluasi sensitivitas kedua skema yang ditunjukkan pada Gambar 5, dapat dilihat bahwa hanya metode *Decision Tree* saja yang mengalami peningkatan yang sebelumnya bernilai 92% menjadi 94%. Sementara metode lainnya tidak mengalami perubahan nilai baik di skema dengan 18 atribut maupun 9 atribut. Hal ini menunjukkan bahwa reduksi jumlah atribut menjadi 50% dari total atribut pada dataset *Cervical Cancer Risk Behavior* tidak begitu mempengaruhi kemampuan model dalam menentukan data berlabel “Bukan Kanker Serviks” diantara data-data yang benar-benar berlabel “Bukan Kanker Serviks”.



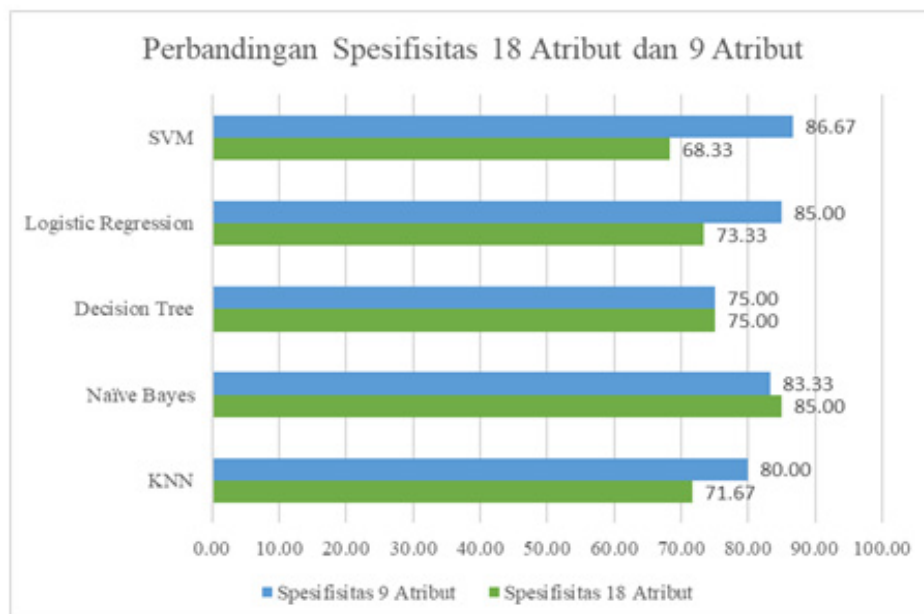
Sumber: Data Primer Diolah, 2022

Gambar 4
Perbandingan Hasil Akurasi dari Skema 18 Atribut dan 9 Atribut



Sumber: Data Primer Diolah, 2022

Gambar 5
Perbandingan Hasil Sensitivitas dari Skema 18 Atribut dan 9 Atribut



Sumber: Data Primer Diolah, 2022

Gambar 6
Perbandingan Hasil Spesifisitas dari Skema 18 Atribut dan 9 Atribut

Pada Gambar 6 dapat diketahui hasil dari perbandingan nilai spesifisitas antara skema dengan 18 atribut dan 9 atribut. Dari hasil tersebut, dapat dilihat hampir semua metode mengalami kenaikan performa spesifisitas.

Kenaikan tertinggi dialami metode SVM, dimana nilai spesifisitas saat menggunakan skema 18 atribut adalah 68,33%, sedangkan nilai tersebut meningkat menjadi 86,67% ketika menggunakan skema 9 atribut.

Pencapaian tersebut diikuti dengan metode Logistic Regression dan KNN yang dimana masing-masing metode ketika menggunakan skema 18 atribut memperoleh nilai spesifisitas sebesar 73,33% dan 71,67% yang kemudian meningkat menjadi 85% dan 80%. Hal berbeda terjadi pada nilai spesifisitas metode Naïve Bayes, dimana hasil skema 9 atribut menurun dibandingkan skema 18 atribut. Hal ini disebabkan oleh perubahan jumlah komponen data ketika menggunakan skema 9 atribut yang dimana metode Naïve Bayes kesulitan dalam mengidentifikasi data dengan label “Kanker Serviks”.

Secara keseluruhan, penggunaan teknik *information gain* sebagai bagian dari analisa multivariat dan sebagai penyeleksi fitur dapat meningkatkan performa dari metode-metode klasifikasi yang digunakan dalam pengujian. Selain itu, skema 9 atribut yang diusulkan mampu memberikan peningkatan performa secara keseluruhan dibandingkan dengan skema 18 atribut. Dengan hasil tersebut diharapkan skema dan model yang diusulkan dapat menjadi acuan dalam mengembangkan sistem deteksi dini kanker serviks berdasarkan data perilaku berisiko.

SIMPULAN

Evaluasi prediksi kanker serviks berdasarkan data perilaku berisiko telah dilakukan. Dari hasil evaluasi menggunakan dataset *Cervical Cancer Risk Behavior* yang terdiri dari 19 atribut utama dan 72 data, didapatkan bahwa teknik *information gain* mampu memberikan rekomendasi fitur berdasarkan bobot terbesar yang relevan terhadap kemunculan kanker serviks pada pasien. Penggunaan skema 9 atribut (50% dari keseluruhan atribut) dalam pengujian menggunakan metode *cross validation* mampu memberikan peningkatan performa pada beberapa metode klasifikasi. Secara keseluruhan, metode *Naive Bayes* memberikan performa terbaik dengan pencapaian 93,21% akurasi, 96% sensitivitas, dan 83,33% spesifisitas. Selain itu, peningkatan performa secara drastis dialami oleh metode SVM. Sehingga dapat disimpulkan bahwa penggunaan skema 9 atribut berdasarkan bobot *information gain* mampu memberikan

peningkatan kemampuan model data mining dalam mengidentifikasi kemunculan kanker serviks.

PUSTAKA ACUAN

Agustyawati, D. N., Fauzi, H., dan Pratondo, A. (2021) ‘Perancangan Aplikasi Deteksi Kanker Serviks Menggunakan Metode Convolutional Neural Network’. *EProceedings of Engineering*, 8(4), pp. 3908–3924.

Alelyani, S. (2021) ‘Stable Bagging Feature Selection on Medical Data’. *Journal of Big Data*, 8(11), pp. 1-18. <https://doi.org/10.1186/s40537-020-00385-8>

Alpan, K. (2021) ‘Performance Evaluation of Classification Algorithms for Early Detection of Behavior Determinant Based Cervical Cancer’. *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 706–710. <https://doi.org/10.1109/ISMSIT52890.2021.9604718>

Bates, S., Hastie, T., dan Tibshirani, R. (2021) ‘Cross-validation: What Does It Estimate and How Well does it do it?’. *Methodology (stat.ME)*. <https://arxiv.org/pdf/2104.00673.pdf>.

Fransisca, Y. M., dan Adhistry, K. (2021) ‘Analisis Dukungan Keluarga dalam Menangani Permasalahan pada Pasien Kanker Serviks’. *Proceeding Seminar Nasional Keperawatan*, 7(1), pp. 116–123.

Lu, W., Chen, T., Yao, Y., dan Chen, P. (2021) ‘Prevalence of High-Risk Human Papillomavirus and Cervical Lesion Risk Factors: A Population-Based Study in Zhejiang, China 2010–2019’. *Journal of Medical Virology*, 93(8), pp. 5118–5125. <https://doi.org/10.1002/jmv.27034>.

- Riani, E. N., dan Ambarwati, D. (2020) 'Early Detection Kanker Serviks sebagai Upaya Peningkatan Derajat Hidup Perempuan'. *SELAPARANG Jurnal Pengabdian Masyarakat Berkemajuan*, 3(2), pp. 144-146. <https://doi.org/10.31764/jpmb.v3i2.1883>.
- Sarosa, S. J. A., Utaminigrum, F., dan Bachtiar, F. A. (2019) 'Breast Cancer Classification Using GLCM and BPNN'. *International Journal of Advances in Soft Computing and Its Applications*, 11(3), pp. 157-172.
- Shalikhah, S., Santoso, S., dan Widyasih, H. (2021) 'Dukungan Keluarga dan Perilaku Deteksi Dini Kanker Serviks Pada Wanita Usia Subur'. *Jurnal Ilmiah Kebidanan*, 9(1), pp. 1-7. <https://doi.org/https://doi.org/10.33992/jik.v9i1.1472>.
- Sobar, Machmud, R., dan Wijaya, A. (2016) 'Behavior Determinant Based Cervical Cancer Early Detection with Machine Learning Algorithm'. *Advanced Science Letters*, 22(10), pp. 3120-3123. <https://doi.org/10.1166/asl.2016.7980>.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., dan Bray, F. (2021) 'Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries'. *CA: A Cancer Journal for Clinicians*, 71(3), pp. 209-249. <https://doi.org/10.3322/caac.21660>.