

KOMPARASI ALGORITMA LDA DAN NAÏVE BAYES DENGAN OPTIMASI FITUR UNTUK KLASIFIKASI CITRA TUNGGAL PAP SMEAR

Yudi Ramdhani

Universitas BSI Bandung, yudi.yrm@bsi.ac.id

ABSTRACT - Cervical cancer is the growth of a group of abnormal cells on the cervix (cervical). According to the WHO (2003), Indonesia is a country with cervical cancer the number one in the world. Algorithm applied in image classification algorithm using the Pap Smear Linear Discriminant Analysis (LDA) and Naïve Bayes algorithm, Genetic Algorithm uses feature selection. The results obtained using the LDA algorithm and Naïve Bayes algorithm do comparisons, the results obtained that the LDA algorithm better than the Naïve Bayes algorithm for image classification is a Pap Smear. The results of the classification into 7 classes have the highest value 62,92% while for the highest classification in the classification of the image of a Pap Smear with 95,87% accuracy rating, there are still difficulties for the classification of Abnormal (4, 5, 6, 7) and the value of accuracy 60,15%. The low accuracy against abnormal class affects the classification into 7 classes.

Key Words: Cervical Cancer, Pap Smears, Linear Discriminant Analysis (LDA), Naïve Bayes, Genetic Algorithm.

ABSTRAKSI - Kanker serviks merupakan pertumbuhan dari suatu kelompok sel yang tidak normal pada serviks (mulut rahim). Menurut WHO (2003) Indonesia merupakan negara dengan penderita kanker mulut rahim nomor satu di dunia. Algoritma yang diterapkan dalam klasifikasi citra *Pap Smear* menggunakan algoritma *Linear Discriminant Analysis* (LDA) dan algoritma *Naïve Bayes*, seleksi fitur menggunakan *Genetic Algorithm*. Hasil yang diperoleh menggunakan algoritma LDA dan algoritma Naïve Bayes dilakukan komparasi, hasil yang diperoleh bahwa algoritma LDA lebih baik dibandingkan dengan algoritma Naïve Bayes untuk klasifikasi citra *Pap Smear*. Hasil klasifikasi ke dalam 7 kelas memiliki nilai tertinggi 62.92% sedangkan untuk klasifikasi tertinggi pada klasifikasi citra *Pap Smear* dengan nilai akurasi 95.87% tetapi masih terdapat kesulitan untuk klasifikasi Abnormal (4,5,6,7) dengan nilai akurasi 60.15%. Akurasi yang rendah terhadap kelas abnormal mempengaruhi klasifikasi kedalam 7 kelas.

Kata Kunci: Kanker Serviks, *Pap Smear*, *Linear Discriminant Analysis* (LDA), *Naïve Bayes*, *Genetic Algorithm*.

I. PENDAHULUAN

Kanker serviks merupakan pertumbuhan dari suatu kelompok sel yang tidak normal pada serviks (mulut rahim). Menurut WHO (2003) Indonesia merupakan negara dengan penderita kanker mulut rahim nomor satu di dunia (Intansari, Purnami, & Wulandari, 2012). Di Indonesia, prevalensi penyakit kanker juga cukup tinggi. Berdasarkan data Riset Kesehatan Dasar tahun 2013, kanker

yang paling sering menyerang perempuan adalah kanker payudara dan kanker serviks dengan 5.349 kasus (12,8%) untuk kanker serviks (Dewi & Purnami, 2015).

Pencegahan penyakit ini dapat dilakukan dengan deteksi dini melalui tes pap smear. Keluaran dari tes pap smear adalah citra sel serviks. Dari citra sel serviks yang dihasilkan, dapat ditentukan sel serviks tersebut masuk dalam kategori

normal atau kanker. Deteksi sel serviks sebagai sel normal atau sel kanker dapat dilakukan melalui dua cara, yakni melalui pengamatan visual, maupun dengan bantuan komputer. Pada umumnya, deteksi kanker serviks dilakukan oleh dokter melalui pengamatan visual (Dharmawan, 2014). Suatu prosedur pap smear konvensional yang dilakukan dengan sangat baik dapat menghasilkan akurasi hasil sebesar 76% sampai dengan 94% (Andrijono, H. Nurwijaya, & H. Suheimi, 2010). Oleh karena itu diperlukan suatu penelitian untuk mendapatkan suatu model yang mampu mengklasifikasikan hasil pap smear dengan tingkat kesalahan klasifikasi yang minimal (Dewi & Purnami, 2015).

Data mining dapat diaplikasikan di bidang kesehatan misalnya mendiagnosis penyakit kanker payudara, penyakit jantung, penyakit diabetes dan lain-lain (Larose, 2005). Kesukaran klasifikasi untuk tujuh kelas dimungkinkan karena jumlah 20 fitur yang memiliki kombinasi yang cukup beragam dalam sebaran tujuh kategori. (Zhang & Srihari, 2002). Algoritma genetika merupakan salah satu algoritma optimasi seleksi fitur. salah satu proses seleksinya adalah dengan mengambil beberapa individu terbaik. selain itu juga dapat dilakukan dengan proses pengambilan acak proporsional, dengan proporsi setara dengan proporsi kualitasnya. (Sartono, 2010). *Linear Discriminant Analysis* merupakan sebuah metode dalam mengekstraksi ciri suatu objek. Dalam mengekstraksi ciri, LDA membagi beberapa kelas dalam pencarian ciri, LDA memiliki ketepatan yang sangat tinggi dalam mengenali wajah seseorang (Brilian, Huda, & Taufik, 2014).

II. KAJIAN LITERATUR

Menggunakan metode *Naïve Bayes* untuk mengidentifikasi sel normal dan abnormal serta mengelompokkannya ke dalam tujuh kelas yang lebih spesifik. Metode klasifikasi *Naïve Bayes* memberikan hasil yang memuaskan pada klasifikasi dua kelas target, tetapi metode

ini belum memberikan hasil yang baik untuk klasifikasi tujuh kelas target (Auliani, 2014). Dwiza Riana pada tahun 2010 dalam penelitiannya model klasifikasi multi tahap dengan memanfaatkan analisa *Importance Performance Analysis* sebagai basis klasifikasi baru yang dapat melakukan proses klasifikasi menjadi 7 kelas citra tunggal *Pap Smear* dengan menggunakan data 210 citra *Pap Smear* dengan sample 30 disetiap kelas dengan tingkat akurasi mencapai 86,71% dengan empat *rule independen*. Mansoori, Suman, Mishra (2014) dalam penelitiannya, algoritma yang digunakan untuk klasifikasi sel kanker dengan menerapkan *Genetic Algorithm* untuk seleksi fitur dan kemudian mengklasifikasikan sel-sel sehat dan sel-sel kanker menggunakan algoritma SVM (*Support Vector Machine*). *Genetic Algorithm* adalah metode yang lebih baik untuk seleksi fitur dan optimasi parameter. Fitur yang terbaik yang dipilih digunakan oleh SVM digunakan untuk klasifikasi pada dataset *training* untuk mengklasifikasikan sel. Hasilnya menunjukkan efektivitas kerja metode ini adalah memberikan akurasi yang tinggi untuk memilih 20 fitur.

A. Kanker Serviks

Kanker serviks adalah tidak terkontrolnya pertumbuhan siklus sel yang ada pada leher rahim (abnormal). Kanker serviks disebut juga kanker leher rahim atau kanker mulut rahim mulai tumbuh pada lapisan serviks. Kanker serviks terbentuk sangat perlahan dan sangat sulit terdeteksi di awal kecuali terjadi infeksi pada fisik. Pertama, tumbuhnya siklus sel kanker yang ada dalam tubuh berubah dari keadaan normal menjadi sekumpulan sel-sel prakanker yang kemudian berkembang menjadi sel kanker. Keadaan perubahan sel kanker ini terjadi secara kontinu, bertahap, dan memerlukan waktu yang bertahun-tahun, atau bahkan tidak jarang pertumbuhan siklus sel kanker ini berlangsung dalam kurun waktu yang relatif cepat. Hal ini disebabkan karena sistem kerja imunitas (daya tahan tubuh) setiap orang berubah-ubah. Perubahan ini

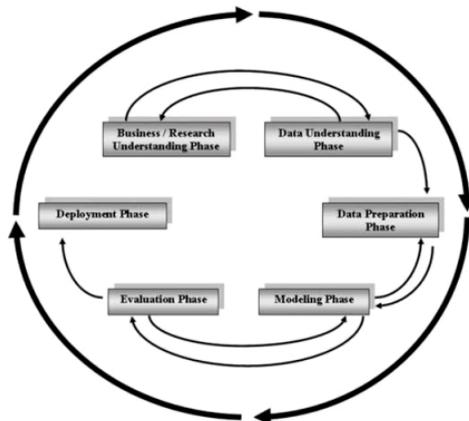
sering disebut displasia. Mereka dapat ditemukan dengan tes Pap Smear dan dapat diobati untuk mencegah terjadinya kanker (Ramdani & Purnami, 2014).

B. Data Mining

Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi data mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu (Witten & Frank, 2011). Data mining merupakan bagian dari *Knowledge Discovery Data* (KDD) yang merupakan proses ekstraksi informasi yang berguna, tidak diketahui sebelumnya, dan tersembunyi dari data (Bramer, 2013) dan juga mengembangkan model yang digunakan untuk memahami fenomena dari analisis data dan prediksi (Maimon & Rokach, 2010).

C. Metodologi Penelitian Data Mining

Metode data mining dan model melanjutkan cakupan data mining sebagai suatu proses, standar proses yang digunakan adalah kerangka kerja CRISP-DM. Dalam CRISP-DM data mining dilihat sebagai proses keseluruhan dari komunikasi masalah bisnis hingga penerapan model (Larose, 2006).



Gambar 2.3. Proses CRISP-DM
Sumber : Larose, 2006

Pada gambar 2.3 Dapat dilihat bahwa proses CRISP-DM terdiri dari enam tahap, yaitu (Larose, 2006):

1. Tahap *Business Understanding*

Tahap *business understanding* disebut juga sebagai tahap *research understanding*, dalam tahap ini ditentukan tujuan dan *requirement* secara detail pada keseluruhan penelitian, merumuskan masalah data mining, dan menyiapkan strategi awal untuk mencapai tujuan yang telah ditentukan.

2. Tahap *Data Understanding*

Pada tahap ini mulai dilakukan proses pengumpulan data, menganalisis data, mengevaluasi kualitas data, dan memilih subset yang mungkin mengandung pola yang ditindaklanjuti.

3. Tahap *Data Preparation*

Pada tahap ini data akhir yang akan digunakan pada tahap berikutnya mulai disiapkan, memilih kasus dan variabel yang sesuai dengan analisis yang akan dilakukan, melakukan transformasi pada variabel tertentu jika diperlukan, dan membersihkan data mentah sehingga siap untuk digunakan sebagai alat pemodelan.

D. Feature Selection

Seleksi fitur adalah terkait erat dengan masalah pengurangan dimensi dimana tujuannya adalah untuk mengidentifikasi fitur dalam kumpulan data-sama pentingnya, dan membuang fitur lain seperti informasi yang tidak relevan dan berlebihan dan akurasi dari seleksinya pada masa depan dapat ditingkatkan (Maimon & Rokach, 2010). Seleksi fitur adalah salah satu faktor yang paling penting yang dapat mempengaruhi tingkat akurasi klasifikasi karena jika dataset berisi sejumlah fitur, dimensi ruang akan menjadi besar dan non-bersih, merendahkan tingkat akurasi klasifikasi (Liu, Wang, Chen, Dong, Zhu, & Wang, 2011).

E. Genetic Algorithms (GA)

Algoritma Genetika adalah algoritma yang digunakan untuk proses pencarian dan optimasi yang didasarkan pada prinsip genetika dan proses seleksi secara *natural*. Algoritma genetika membuat populasi yang tersusun dari

banyak individu yang berevolusi menurut aturan seleksi tertentu yang mempunyai ketetapan optimasi dan nilai. (Weise, 2009).

F. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) pertama kali diterapkan pada proses pengenalan wajah oleh Etemad dan Chellapa. LDA bekerja berdasarkan analisa matrik penyebaran (*scatter matrix analysis*) yang bertujuan menemukan suatu proyeksi optimal yang dapat memaksimalkan penyebaran antar kelas dan meminimumkan penyebaran dalam kelas data wajah. Algoritma LDA memiliki karakteristik perhitungan matriks yang hampir sama dengan PCA. Perbedaan dasarnya adalah pada LDA diusahakan adanya perbedaan yang minimum dari citra dalam kelas. Perbedaan antar kelas direpresentasikan oleh matriks S_b (*scatter between class*) dan perbedaan dalam kelas direpresentasikan oleh matriks S_w (*scatter within class*). Matriks *covariance* didapatkan dari kedua matriks tersebut. Untuk memaksimalkan jarak antar kelas dan meminimumkan jarak dalam kelas digunakan suatu *discriminant power* (Kusuma & Wirawan, 2008).

G. Naïve Bayes

Naive Bayes merupakan metode yang tidak memiliki aturan (Alfisahrin, 2014). *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada *data training*. *Naive Bayes* merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam *data mining* (Wu & Kumar, 2009). Algoritma ini juga dikenal dengan nama *Idiot's Bayes*, *Simple Bayes*, dan *Independence Bayes* (Bramer, 2013).

Klasifikasi *Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Klasifikasi Bayesian didasarkan pada teorema Bayes, diambil dari nama seorang ahli matematika yang juga

menteri *Prebysterian* Inggris, Thomas Bayes (1702-1761) (Bramer, 2013). Klasifikasi bayesian memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* (Alfisahrin, 2014).

III. METODE PENELITIAN

A. Desain Penelitian

Pengertian penelitian dalam akademik yaitu digunakan untuk mengacu pada aktivitas yang rajin dan menyelidiki sistematis atau investigasi di suatu daerah, dengan tujuan menemukan atau merevisi fakta, teori, aplikasi dan tujuannya adalah untuk menemukan dan menyebarkan pengetahuan baru (Berndtsson, Olsson, & Lundell, 2008).

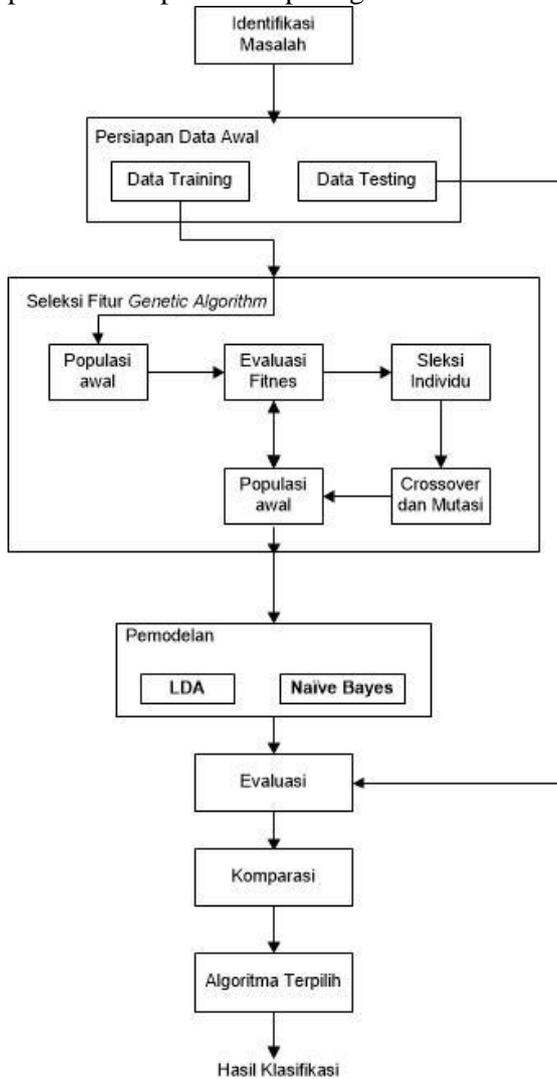
B. Pengumpulan Data

Data yang digunakan untuk *training* dan *testing* merupakan data *public* yang diklasifikasikan dengan hati-hati oleh *cyto-technicians* dan dokter. Untuk meningkatkan klasifikasi citra sel *Pap Smear* dalam eksperimen ini digunakan data Herlev 917 (Samuel, Mathew, dan Sreekumar, 2014). Citra *Pap Smear* diklasifikasikan kedalam 7 kelas yaitu: *Normal Superficial*, *Normal Intermediat*, *Normal Colummar*, *Mild (Light) Dyplasia*, *Moderate Dyplasia*, *Severe Dyplasia* dan *Carcinoma In Situ* dengan citra sebanyak 917. Data diperoleh dari Herlev University Hospital, Denmark yang diperoleh melalui halaman *website* : <http://labs.fme.aegean.gr/decision/download>.

C. Metode yang Diusulkan

Dalam penelitian ini akan memilih dan menerapkan teknik yang tepat untuk klasifikasi citra *Pap Smear*. Tahap pertama dalam penelitian ini adalah membagi *dataset* citra sel *Pap Smear* dengan membagi dua yaitu, data *training* dan data *testing*. Langkah selanjutnya adalah melakukan seleksi fitur terbaik dalam *dataset* citra *Pap Smear* menggunakan *Genetic Algorithms* (GA). Kemudian fitur yang telah diseleksi dilakukan diklasifikasi menggunakan algoritma *Linear Discriminant Analysis*

(LDA) dan *Naïve Bayes* , model terbaik dari hasil klasifikasi dibandingkan untuk melihat model yang terbaik dari hasil klasifikasi. Hasil dari klasifikasi akan diukur dengan nilai akurasi. Desain penelitian dapat dilihat pada gambar 3.1.



Gambar 3.1 Desain Penelitian

IV. PEMBAHASAN

Dalam penelitian ini akan dilakukan eksperimen seleksi fitur menggunakan *Genetic Algorithm* dan klasifikasi *Pap Smear* menggunakan algoritma *LDA* dan *Naïve Bayes* .

Hasil penelitian klasifikasi citra *Pap Smear* menggunakan algoritma *LDA* dan *Naïve Bayes* telah dilakukan, hasil perbandingan masing-masing klasifikasi dapat dilihat pada tabel 4.1.

Tabel 4.1 Hasil Klasifikasi Algoritma *LDA* dan *Naïve Bayes*

No	Klasifikasi	Algoritma	
		LDA	<i>Naïve Bayes</i>
1	7 Kelas	59.21%	55.51%
3	Normal dan Abnormal	90.08%	91.60%
2	Normal 1,2,3	93.80%	93.39%
4	Kelas Abnormal 4, 5, 6, 7	58.07%	50.81%

Setelah melakukan analisa terhadap hasil eksperimen maka diperoleh nilai tertinggi sebesar 93.80% dengan klasifikasi kelas Normal (1,2,3). Algoritma *Naïve Bayes* mengguguli algoritma *LDA* dengan hasil klasifikasi 91.60% dengan klasifikasi kelas Normal dan Abnormal. Sedangkan klasifikasi dengan 7 kelas memiliki akurasi terbesar 59.21% memiliki akurasi yang kurang baik untuk klasifikasi citra *Pap Smear* untuk itu dilakukan perbaikan akurasi pada langkah berikutnya.

Untuk meningkatkan hasil klasifikasi maka dilakukan seleksi fitur menggunakan *Genetic Algorithm* (GA). Hasil dari klasifikasi citra *Pap Smear* menggunakan algoritma *LDA* dan *Naïve Bayes* dengan seleksi fitur menggunakan *Genetic Algorithm* dapat dilihat dari table 4.2.

Tabel 4.2 Hasil klasifikasi Algoritma LDA dan *Naïve Bayes* dengan GA

No	Klasifikasi	Algoritma	
		LDA + GA	<i>Naïve Bayes</i> + GA
1	7 Kelas	62.92%	62.16%
3	Normal dan Abnormal	91.06%	93.02%
2	Normal 1,2,3	95.87%	95.04%
4	Kelas Abnormal 4, 5, 6, 7	60.15%	57.33%

Setelah melakukan analisa terhadap hasil eksperimen maka diperoleh nilai tertinggi dari klasifikasi sebesar 95.87% dengan klasifikasi kelas Normal (1,2,3). Algoritma *Naïve Bayes* menungguli algoritma LDA dengan nilai klasifikasi 93,02% dengan klasifikasi kelas Normal dan Abnormal. Sedangkan untuk klasifikasi 7 kelas sebesar 62.92% menggunakan algoritma LDA.

Setelah melakukan eksperimen menggunakan algoritma LDA dan *Naïve Bayes* untuk klasifikasi citra *Pap Smear* dan algoritma seleksi fitur menggunakan GA maka dilakukan komparasi hasil klasifikasi dengan hasil yang diperoleh pada table 4.3.

Tabel 4.3 Hasil Komparasi Algoritma LDA dan *Naïve Bayes*

Klasifikasi	Algoritma			
	LDA + GA	LDA	<i>Naïve Bayes</i> + GA	<i>Naïve Bayes</i>
7 Kelas	62.92 %	59.2 %	62.16 %	55.51 %
Normal dan Abnormal	91.06 %	90.0 %	93.02 %	91.60 %
Normal 1,2,3	95.87 %	93.8 %	95.04 %	93.39 %
Kelas Abnormal 4, 5, 6, 7	60.15 %	58.0 %	57.33 %	50.81 %

Hasil yang diperoleh setelah melakukan klasifikasi menggunakan algoritma klasifikasi LDA dan *Naïve Bayes* dengan menggunakan seleksi fitur GA dan tanpa menggunakan seleksi fitur GA dapat terlihat bahwa menggunakan seleksi fitur menggunakan GA meningkatkan hasil akurasi dengan nilai sebesar 62.92% untuk klasifikasi 7 kelas,

sedangkan dengan menggunakan algoritma LDA tanpa menggunakan seleksi fitur GA memiliki nilai sebesar 59.21%. Nilai tertinggi dari hasil klasifikasi adalah 95.87% untuk klasifikasi normal (1,2,3).

V. PENUTUP

A. Kesimpulan

Dalam penelitian ini telah dilakukan pengujian model klasifikasi citra *Pap Smear* dengan menggunakan algoritma LDA dan *Naïve Bayes* da seleksi fitur menggunakan *Genetic Algorithm* dengan hasil penelitian sebagai berikut:

1. Model klasifikasi 7 kelas (*Normal Superficial, Normal Intermediat, Normal Colummar, Mild (Light) Dyplasia, Moderate Dyplasia, Severe Dyplasia dan Carcinoma In Situ*) menggunakan algoritma LDA dengan nilai akurasi tertinggi 62.92%.
2. Model klasifikasi Normal (*Normal Superficial, Normal Intermediat, Normal Colummar*) dan Abnormal (*Mild (Light) Dyplasia, Moderate Dyplasia, Severe Dyplasia dan Carcinoma In Situ*) menggunakan algoritma klasifikasi *Naïve Bayes* dan seleksi fitur menggunakan *Genetic Algorithm* dengan nilai akurasi tertinggi 93.02%.
3. Model klasifikasi Normal (*Normal Superficial, Normal Intermediat, Normal Colummar*) menggunakan algoritma klasifikasi LDA dan seleksi fitur menggunakan *Genetic Algorithm* dengan nilai akurasi tertinggi 95.87%.
4. Model klasifikasi Abnormal 4, 5, 6, 7 (*Mild (Light) Dyplasia, Moderate Dyplasia, Severe Dyplasia dan Carcinoma In Situ*) menggunakan algoritma klasifikasi LDA dan seleksi fitur menggunakan *Genetic Algorithm* dengan nilai akurasi tertinggi 60.15%.

Hasil yang diperoleh dari eksperimen menunjukkan bahwa klasifikasi tertinggi terhadap klasifikasi Normal (1,2,3) dengan akurasi 95.87%. tetapi masih terdapat kesulitan untuk klasifikasi Abnormal (4,5,6,7) dengan

nilai akurasi 60.15%. Akurasi yang rendah terhadap kelas abnormal mempengaruhi klasifikasi kedalam 7 kelas.

Algoritma LDA memiliki akurasi yang baik untuk mengklasifikasikan data kedalam Polynominal atau memiliki klasifikasi lebih dari 2 pilihan, sedangkan *Naïve Bayes* memiliki akurasi yang baik terhadap klasifikasi Binominal atau klasifikasi terhadap 2 pilihan.

B. Saran

Diketahui bahwa algoritma klasifikasi LDA memiliki akurasi yang tinggi untuk klasifikasi citra *Pap Smear*, namun untuk penelitian selanjutnya hal-hal berikut dapat ditambahkan untuk meningkatkan akurasi dan efektifitas klasifikasi citra *Pap Smear* dapat dilakukan, diantaranya:

1. Menggunakan metode optimasi bobot, optimasi parameter, dan seleksi fitur dengan menggunakan algoritma lainnya seperti seperti *Particle Swarm Optimimization*, *Greedy* dan lain-lainnya.
2. Menggunakan algoritma klasifikasi lainnya untuk meningkatkan klasifikasi citra *Pap Smear*.

REFERENCE

- Alfisahrin, S. N. (2014). *Komparasi Algoritma C4.5, Naive Bayes dan Neural Network Untuk Memprediksi Penyakit Jantung*. Jakarta: Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
- Andrijono, H. Nurwijaya, & H. Suheimi. (2010). *Cegah dan Deteksi Kanker*. Jakarta: Elex Media.
- Auliani, D. L. (2014). Deteksi Kanker Serviks Berdasarkan Citra Sel *Pap Smear* Dengan Klasifikasi *Naïve Bayes*.
- Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2008). *A Guide For Students In Computer Science And Information Systems*. London: Springer.
- Bramer, M. (2013). *Principle of Data Mining Second Edition*. London: Springer.
- Brilian, F., Huda, A. F., & Taufik, I. (2014). Sistem Pengenalan Wajah Dengan Menggunakan Metode Linear Discriminant Analysis (LDA).
- Dewi, M. R., & Purnami, S. W. (2015). Klasifikasi Hasil *Pap Smear Test* sebagai Upaya Pencegahan Sekunder Penyakit Kanker Serviks di Rumah Sakit "X" Surabaya Menggunakan Piecewise Polynomial Smooth Support Vector Machine (PPSSVM). *JURNAL SAINS DAN SENI ITS*, 61-66.
- Dharmawan, D. A. (2014). Deteksi Kanker Serviks Otomatis Berbasis Jaringan Saraf Tiruan LVQ dan DCT. *JNTETI*, 269-272.
- Faried, A. (2012). Bagaimana mereka (sel kanker) berjalan? 1-6.
- Intansari, I. A., Purnami, S. W., & Wulandari, S. P. (2012). Klasifikasi Pasien Hasil *Pap Smear Test* sebagai Pendeteksi Awal Upaya Penanganan Dini pada Penyakit Kanker Serviks di RS. "X" Surabaya dengan Metode Bagging Logistic Regression. *JURNAL SAINS DAN SENI ITS*, 277-282.
- Kusuma, H., & Wirawan. (2008). Teknik Pengenalan Wajah dengan Metoda Subspace LDA (Linear Discriminant Analysis).
- Larose, D. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. United States of America: John Wiley & Sons, Inc.
- Larose, D. (2006). *Data Mining Methods And Models*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook Second Edition*. New York: Springer.
- Mansoori, T. K., Suman, A., & Mishra, S. K. (2014). Feature Selection by Genetic Algorithm and SVM Classification for Cancer Detection. *International Journal of Advanced Research in Computer Science and Software Engineering Vol 4*, 357-365.
- Mustafa, N., Isa, N. M., Mashor, M., & Othman, N. (2007). New Features of Cervical Cells for Cervical Cancer Diagnostic System Using Neural Network.
- Ramdani, Y., & Purnami, S. W. (2014). Klasifikasi Hasil *Pap Smear Test*

- Kanker Serviks Berdasarkan Faktor Resiko (Studi Kasus Di Rumah Sakit Swasta Surabaya). 1-6.
- Riana, D. (2010). *Hierarchical Decision Approach Berdasarkan Importance Performance Analysis Untuk Klasifikasi Citra Tunggal Pap Smear Menggunakan Fitur Kuantitatif dan Kualitatif*. Depok : Universitas Indonesia: Tesis Tidak Terpublikasi.
- Samuel, S. S., Mathew, A. V., & Sreekumar, S. (2014). Comparative Study Between Sparse Representation Classification and Classical Classifiers on Cervical Cancer Cell Images. *International Journal of Advanced Research in Computer and Communication Engineering*.
- Sartono, B. (2010). Pengenalan Algoritma Genetik Untuk Pemilihan Peubah Penjelas Dalam Model Regresi Menggunakan SAS/IML. *Forum Statistika dan Komputasi*, 10-15.
- Witten, I., & Frank, E. (2011). *Data Mining Practical Machine Learning Tools and Techniques 3rd Edition*. Burlington, Usa: Morgan Kaufmann Publishers.
- Wu, X., & Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. New York: CRC Press.
- Zhang, B., & Srihari, S. N. (2002). Class-Wise Multi-Classifer Combination Based On Dempster-Shafer Theory. *In Proceedings The Seventh International Conference On Control, Automation, Robotics And Vision*.