

Implementasi Algoritma PSO Dan Teknik *Bagging* Untuk Klasifikasi Sel *Pap Smear*

Toni Arifin

Universitas BSI
e-mail: toni.tfn@bsi.ac.id

Abstrak

Penyakit kanker adalah salah satu penyebab kematian di seluruh dunia. Di Indonesia Kanker serviks dan kanker payudara merupakan penyakit kanker dengan penderita terbanyak. Penyebab kanker serviks adalah virus HPV (*Human Papilloma Virus*) tipe 16 dan 18. Tes Pap Smear merupakan salah satu pencegahan kanker serviks secara dini. Pada pemeriksaan Pap Smear sel akan di amati di bawah mikroskop untuk membedakan sel normal dan abnormal, pada pemeriksaan ini ahli patologi terkadang kesulitan dalam pengamatan sel karena bentuk sel yang hampir mirip, dan pemeriksaan sel memakan waktu dan terkadang terjadi kesalahan. Tujuan dari penelitian ini adalah mengusulkan model klasifikasi untuk klasifikasi sel Pap Smear untuk memudahkan ahli patologi. Metode yang digunakan adalah kombinasi dari metode *Particle Swarm Optimization* untuk seleksi fitur dan Teknik *Bagging* untuk mengatasi jumlah kelas yang tidak seimbang. Dari kombinasi ke 2 metode tersebut di ujicoba dengan metode klasifikasi *Decision Tree*, *Naïve Bayes* dan K-NN untuk mengetahui perbandingan dari setiap metode klasifikasi. Hasil dari penelitian ini menunjukkan bahwa penggabungan metode *Particle Swarm Optimization* dan Teknik *Bagging* terbukti efektif untuk klasifikasi sel Pap Smear, itu di lihat dari hasil akurasi yang ditunjukkan. Klasifikasi dengan metode K-NN menghasilkan akurasi terbaik untuk klasifikasi sel normal dan abnormal yaitu 95,05%, sedangkan metode klasifikasi dengan akurasi terbaik untuk klasifikasi 7 kelas yaitu *Decision Tree* dengan 64,24%.

Kata Kunci: Kanker Serviks, Pap Smear dan Klasifikasi.

Abstract

Cancer is one of the leading causes of death in worldwide. In Indonesia cervical cancer and breast cancer is a cancer disease with most patients. The cause of cervical cancer is HPV virus (Human Papilloma Virus) types 16 and 18. Pap Smear test is one of the prevention of cervical cancer early. On Pap Smear examination the cells will be observed under a microscope to distinguish normal and abnormal cells, on this examination Pathologists sometimes find it difficult to observe cells because of the almost identical cell shape, and examination are time-consuming and sometimes faulty. The purpose of this research is to propose a classification model for Pap Smear to facilitate Pathologists. The method is used a combination of Particle Swarm Optimization for selection feature and Bagging technique to overcome an unbalanced of classes. From the combination of the two methods, we tested the classification method of Decision Tree, Naïve Bayes and K-NN to find out the comparison of each classification method. The result of this research indicate that the incorporation of Particle Swarm Optimization method and Bagging Technique proved effective for classification of Pap Smear cells, it is viewed from the accuracy shown results. Classification with K-NN method gives the best accuracy for normal and abnormal cell classification of 95,05%, while classification method with best accuracy for classification 7 class is Decision Tree with 64,24%.

Keywords: Cervical Cancer, Pap Smear and Classification.

1. Pendahuluan

Penyakit kanker adalah penyakit yang timbul akibat pertumbuhan tidak normal sel jaringan tubuh yang berubah mejadi sel kanker, sedangkan tumor adalah kondisi dimana pertumbuhan sel tidak normal sehingga membentuk suatu lesi atau dalam banyak kasus benjolan di tubuh. (Kemenkes, 2015)

Kanker merupakan salah satu masalah kesehatan di Indonesia. Berdasarkan data riset kesehatan dasar (Riskesdas) 2013, kasus kanker di Indonesia terjadi sebanyak lebih kurang 330.000 orang dengan kasus terbesar adalah kanker serviks atau kanker leher rahim. Sementara itu data dari WHO *information centre on HPV and cervical cancer* menyatakan bahwa 2 dari 10.000 wanita di Indonesia menderita kanker serviks dan di perkirakan 26 wanita meninggal setiap harinya karena kanker serviks.

Penyakit kanker serviks dan payudara merupakan penyakit kanker dengan prevalensi tertinggi di Indonesia pada tahun 2013, yaitu kanker serviks sebesar 0,8% dan kanker payudara sebesar 0,5%. (Kemenkes, 2015)

Kanker serviks merupakan keganasan yang bersal dari serviks. Serviks merupakan sepertiga bagian bawah uterus, berbentuk silindris, menonjol dan berhubungan dengan vagina melalui ostium uteri eksternum. (Kemenkes, 2015)

Pada tahun 2010 estimasi jumlah insiden kanker serviks adalah 454.000 kasus. Data ini didapatkan dari registrasi kanker berdasarkan populasi, registrasi data vital, dan data otopsi verbal dari 187 negara dari tahun 1980 sampai 2010. Per tahun insiden dari kanker serviks meningkat 3.1% dari 378.000 kasus pada tahun 1980. Ditemukan sekitar 200.000 kematian terkait kanker serviks, dan 46.000 diantaranya adalah wanita usia 15-49 tahun yang hidup di negara sedang berkembang. (Kemenkes, 2015)

Kanker serviks sebenarnya dapat di cegah dengan melakukan upaya pencegahan primer seperti meningkatkan atau intensifikasi kegiatan penyuluhan kepada masyarakat untuk menjalankan pola hidup sehat, menghindari faktor resiko terkena kanker, melakukan imunisasi

dengan vaksi HPV dan diikuti dengan deteksi dini kanker serviks melalui pemeriksaan Pap Smear. (Septadina, Kesuma, Handayani, Suciati, & Liana, 2007)

Tes Pap Smear dapat mendeteksi perubahan sel-sel rahim di bawah mikroskop untuk mengetahui apakah ada kemungkinan sel serviks akan menjadi kanker di masa yang akan datang. Mendeteksi sel Abnormal sejak awal dengan melakukan tes Pap Smear, merupakan langkah awal yang tepat dalam menghentikan kemungkinan berkembangnya kanker serviks.

Penelitian ini bertujuan untuk mengklasifikasikan sel Pap Smear dengan menggunakan Teknik data mining. Data yang digunakan pada penelitian ini adalah data dari Pap Smear *Benchmark Data For Pattern Clasification* J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, *University Hospital Dept. of Pathology Harlev Ringvej 75, DK-2730 Harlev Denmark*.

Ada beberapa penelitian yang menjadi acuan peneliti dalam melakukan penelitian mengenai data mining untuk klaifikasi Sel Pap Smear diantaranya adalah penelitian yang dilakukan oleh Martin (2003) melakukan klasifikasi citra sel Pap Smear dengan membaginya ke dalam 2 kelas yaitu kelas normal dan abnormal dan 7 kelas yaitu *Superficial epithelial, Intermediate epithelial, Columnar epithelial, Mild dysplastic, Moderate dysplastic, Severe dysplastic, dan Carcinoma in situ*.

Jantzen, Norup, Dounias & Bjerregaard (2005), melakukan penelitian terhadap 917 sampel citra sel tunggal Pap Smear dan menghasilkan data set dari 917 citra sel tunggal Pap Smear kedalam klasifikasi menjadi 2 kelas yakni normal dan abnormal serta menjadi 7 kelas yakni *Normal Superficial, Normal Intermediate, Normal Columnar, Mild (Light) Dysplasia, Moderate Dysplasia, Severe Dysplasia, dan Carcinoma In Situ*. Data set yang dihasilkan dari penelitian ini dikenal dengan nama data set citra sel tunggal Herlev.

Wahono & Suryana (2013), dalam penelitiannya Algoritma *Particle swarm optimization* diterapkan untuk menangani pemilihan fitur dan teknik bagging digunakan untuk mengatasi ketidakseimbangan kelas. Metode yang diusulkan dievaluasi menggunakan kumpulan data dari data repository NASA. Hasil pada penelitian ini menunjukkan

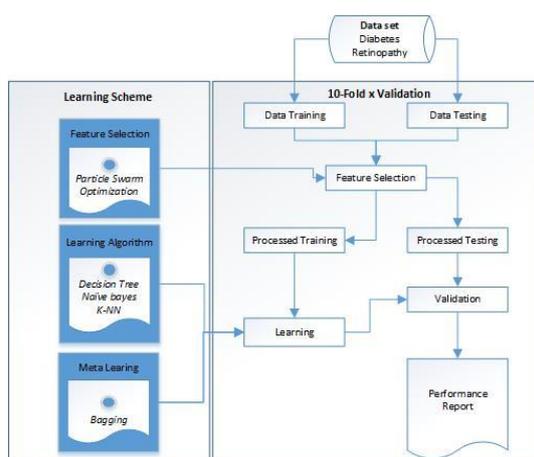
bahwa metode yang diusulkan membuat perbaikan yang mengesankan dalam kinerja prediksi untuk kebanyakan pengklasifikasian.

Penelitian selanjutnya adalah penelitian yang dilakukan oleh Hidayatulloh, Herliana & Arifin (2016). Pada Penelitian ini data yang digunakan adalah data set Sel Pap Smear Herlev Denmark dengan jumlah data adalah 917 data. Metode yang digunakan adalah metode klasifikasi yang memberikan nilai terbaik pada penelitian ini adalah *Naive bayes Classifier*, sedangkan untuk metode optimasi yang digunakan adalah *Particle Swarm Optimization*. Dengan kombinasi metode *Naive bayes* dan *Particle Swarm Optimization*, diperoleh hasil yang lebih baik dari penelitian sebelumnya yaitu 62,67% untuk klasifikasi 7 kelas dan 95,70% untuk klasifikasi 2 kelas.

Pada penelitian ini data yang akan digunakan adalah data set Sel Pap Smear Herlev Denmark yang berjumlah 917 data. Tujuan dari penelitian ini adalah membandingkan metode klasifikasi *Decision Tree*, k-NN dan *Naive Bayes* dengan kombinasi dari 2 metode optimasi yaitu *Particle Swarm Optimization* dan Teknik bagging untuk menghasilkan akurasi yang maksimal.

2. Metode Penelitian

Pada penelitian ini diperlukan kerangka penelitian untuk menggambarkan penelitian yang akan dilakukan, dibawah ini adalah gambar 2.1 kerangka penelitian.



Gambar 2.1 Kerangka Penelitian

Pada kerangka penelitian diatas menjelaskan tentang data set yang akan dibagi menjadi 10 bagian dengan menggunakan *10-Fold x validation*, data

pada bagian pertama menjadi data testing dan ke dua menjadi data training, setelah itu data akan di seleksi dengan menggunakan metode fitur seleksi dengan *Particle Swarm Optimization* dan akan menghasilkan data training dan testing yang terseleksi, setelah itu data training akan diuji dengan metode data mining klasifikasi dan Teknik bagging untuk menangani masalah *imbalance class* lalu di validasi dengan menggunakan *confusion matrix* untuk mengetahui metode klasifikasi mana yang terbaik dengan menggunakan kombinasi dari 2 metode optimasi untuk klasifikasi sel Pap Smear.

Dibawah ini tabel yang menjelaskan jumlah data yang dipakai dan nama atribut yang ada di data set sel Pap Smear.

Tabel 2.1 Nama-nama atribut

Attribute Sel Pap Smear 917	
No	Nama
1	<i>Normal Superficial</i>
2	<i>Normal Intermediate</i>
3	<i>Normal Columnar</i>
4	<i>Mild (Light) Dysplasia</i>
5	<i>Moderate Dysplasia</i>
6	<i>Severe Dysplasia</i>
7	<i>Carcinoma In Situ</i>

Tabel 2.2 Jumlah data 2 kelas Sel Pap Smear

Sel Pap Smear 2 Kelas	
Normal	242
Abnormal	675
Total	917

Tabel 2.3 Jumlah data 7 kelas Sel Pap Smear

Sel Pap Smear 7 kelas	
kelas 1	74
kelas 2	70
kelas 3	98
kelas 4	182
kelas 5	146
kelas 6	197
kelas 7	150
Total	917

a. Particle Swarm Optimization

Particle Swarm Optimization atau sering disebut juga sebagai PSO, merupakan sebuah metode optimasi yang terinspirasi dari gerak dan perilaku hewan seperti ikan dan burung dalam berperilaku seperti mencari mangsa, yang pertama kali dikemukakan oleh James Kennedy dan Russell C. Eberhart pada tahun 1995. PSO terdiri dari sekumpulan partikel yang mencari posisi terbaik, yang merupakan posisi terbaik untuk masalah optimasi dalam ruang fitur. Kelebihan metode optimasi *Particle Swarm Optimization* adalah mempunyai konsep sederhana, mudah diimplementasikan, dan efisien dalam perhitungan jika dibandingkan dengan algoritma matematika dan teknik optimisasi heuristik lainnya (Shih Wei, Kuo-Ching, Shih-Chieh, & Zne-Jung, 2008)

$$V_{i,m} = w \cdot V_{i,m} + c_1 * R * (pbest_{i,m} - x_{i,m}) + c_2 * R * (gbest_m - x_{i,m})$$

$$x_{id} = x_{i,m} + V_{i,m}$$

Keterangan:

- t: Menunjukkan counter iterasi
 V_{ij} : kecepatan partikel i pada dimensi ke j (nilainya terbatas antara $[-v_{max}, v_{max}]$,
 p_{ij} : posisi partikel i pada j dimensi (nilainya terbatas $[-p_{max}, p_{max}]$
 $pbest_{ij}$: posisi pbest partikel i pada dimensi ke-j
 $gbest_{ij}$: posisi gbest dari dimensi ke-j
w: berat inersia (menyeimbangkan eksplorasi global dan lokal eksploitasi)
 $rand_1$ dan $rand_2$: fungsi acak di rentang $[0, 1]$
 β : faktor kendala untuk mengontrol kecepatan berat (nilainya ke 1) c_1 dan c_2 adalah faktor pembelajaran pribadi dan sosial (nilainya ke 2)

b. Bagging

Teknik ensemble merupakan teknik yang sukses untuk menangani dataset yang tidak seimbang meskipun tidak secara khusus dirancang untuk masalah data yang tidak seimbang (Laradji, Alshayeb, & Ghouti, 2015). Teknik bagging merupakan salah satu teknik ensemble dan teknik ini digunakan pada klasifikasi untuk memisahkan data training ke dalam beberapa data training baru dengan random sampling dan membangun model

berbasis data training baru (Wahono & Suryana, 2013). Di bawah ini adalah algoritma Bagging untuk model klasifikasi (Liu & Zhou, 2013).

Input: Data set $D = \{(x_i, y_i)\}_{i=1}^n$

Base learning algorithm \mathcal{L}

The number of iterations T

- 1) for $t=1$ to T do
- 2) $h_t = \mathcal{L}(D, \mathcal{D}_{bs})$
- 3) end for
- 4) Output: $H(x) = \max_y \sum_{t=1}^n I(h_t(x) = y)$

c. Decision tree

Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan yang merupakan pengembangan dari algoritma ID3, algoritma tersebut digunakan untuk membentuk pohon keputusan. Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer, dalam klasifikasi pohon keputusan terdiri dari sebuah node yang membentuk akar, node akar tidak memiliki inputan. Node lain yang bukan sebagai akar tetapi memiliki tepat satu inputan disebut node internal atau test node, sedangkan node lainnya dinamakan daun. Daun mewakili nilai target yang paling tepat dari salah satu class (Maimon & Rokach, 2010). Langkah-langkah membangun pohon keputusan menggunakan algoritma C4.5 menurut Kusriani dan Luthfi adalah sebagai berikut (Arifin, 2014).

- 1) Pilih atribut sebagai akar
Pemilihan atribut sebagai akar berdasarkan pada nilai gain tertinggi dari atribut yang ada. Untuk menghitung nilai gain tertinggi digunakan persamaan

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Nilai Entropy dapat dihitung dengan persamaan di bawah ini:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

- 2) Buat cabang untuk tiap-tiap nilai
- 3) Bagi kasus dalam cabang Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

d. Naïve bayes

Naïve Bayes Classifier atau bias kita sebut dengan NBC ini adalah sebuah metode klasifikasi statistik yang dapat memprediksikan kemungkinan klasifikasi keanggotaan sebuah class (Kusrini & Luthfi, 2009). Dalam pengertian lain, metode *Naïve Bayes Classifier* dikatakan sebagai teknik prediksi berbasis probabilistic sederhana yang berdasar pada penerapan teorema Bayes dengan asumsi independensi yang kuat (Prasetyo, 2012). Kelebihan dari *Naïve Bayes Classifier* adalah dapat digunakan dan dianggap cukup handal untuk mengolah data dan proses klasifikasi yang tidak dapat diselesaikan oleh metode SVM dalam permasalahan jumlah data set yang besar dengan fitur yang beragam dalam arti lain Naïve Bayes Classifier memiliki kelebihan sebagai metode atau algoritma yang sederhana tetapi memiliki akurasi yang tinggi (Rodiansyah & Winarko, 2013). Di bawah ini persamaan dari algoritma Naïve Bayes (Prasetyo, 2012).

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

Keterangan:

$P(H|X)$: Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika diberikan bukti (*evidence*) X terjadi.

$P(X|H)$: Probabilitas sebuah bukti X terjadi akan mempengaruhi hipotesis H.

$P(H)$: Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.

$P(X)$: Probabilitas awal (priori) bukti X terjadi tanpa memandang hipotesis atau bukti yang lain.

e. K-NN

K-Nearest Neighbor termasuk kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *Lazy learning*. *k-nearest neighbor* bekerja dengan mencari kelompok K objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing (Wu, 2008). Algoritma *k-nearest neighbor* merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain. Dekat atau jauhnya lokasi (jarak) biasanya dihitung berdasarkan jarak *Euclidean* atau jarak terdekat (Han &

Kamber, 2006) (Arifin, 2014) Di bawah ini adalah contoh perhitungan Algoritma K-NN Persamaan mencari Tetangga terdekat

$$d_{Euclidean}(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Persamaan normalisasi nilai Min dan Max

$$X^* = \frac{X - \min(X)}{\text{Range}(X)} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

f. Pengujian *Confusion Matrix*

evaluasi kinerja model klasifikasi didasarkan pada pengujian objek yang diprediksi dengan benar dan salah, hitungan ini ditabulasikan *confussion matrix* (Alfisahrin, 2014). *Confussion Matrix* adalah alat yang berguna untuk menganalisis seberapa baik *classifier* dapat mengenali tupel dari kelas yang berbeda (Han & Kamber, 2006).

Tabel 2.4 *Confusion Matrix*

Classification	Predicted Class		
	Class	Yes	No
Observed Class	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

TP: Jumlah kasus positif yang diklasifikasi sebagai positif

FP: Jumlah kasus negatif yang diklasifikasi sebagai positif

TN: Jumlah kasus negatif yang diklasifikasi sebagai negatif

FN: Jumlah kasus positif yang diklasifikasi sebagai negatif

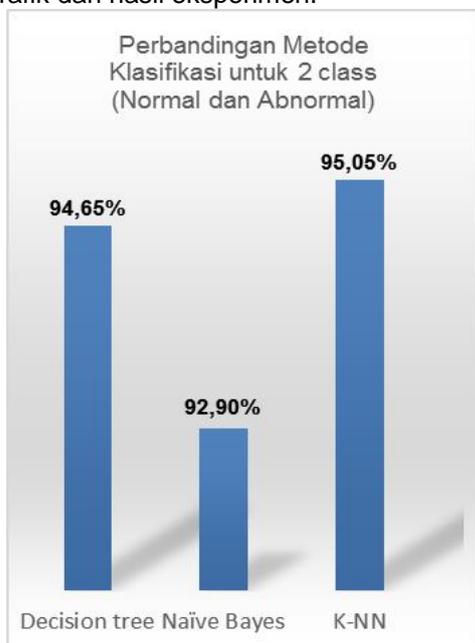
3. Hasil dan Pembahasan

Dari hasil pengujian yang telah dilakukan, hasil dari eksperimen dengan kombinasi dari algoritma *Particle Swarm Optimization* (seleksi fitur) dan algoritma *Bagging (imbalance Class)* beragam di lihat dari akurasi yang di hasilkan dari 3 metode klasifikasi data mining. Di bawah ini adalah tabel 3.1 hasil akurasi yang di hasilkan dari 3 metode klasifikasi dan 2 metode optimasi (*Fiture Selection & Imbalance Class*).

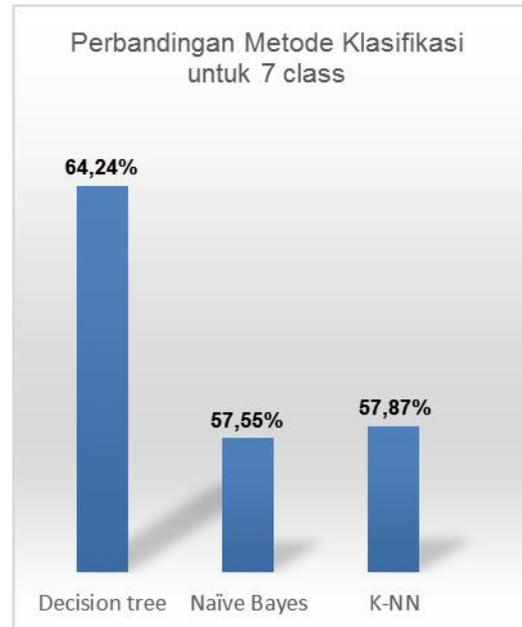
Tabel 3.1 Hasil Pengujian

HASIL PENGUJIAN		
	PSO & Bagging (2 class)	PSO & Bagging (7 class)
Decision tree	94.65%	64.24%
Naïve Bayes	92.90%	57.55%
K-NN	95.05%	57.87%

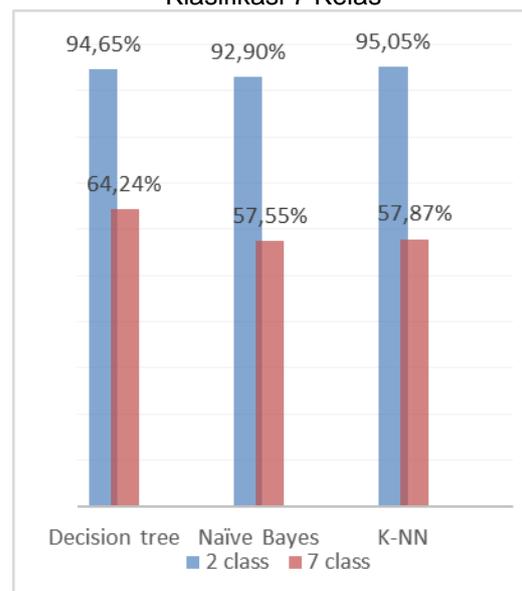
Dari hasil yang ditunjukkan tabel 3.1 di atas perbedaan hasil eksperimen klasifikasi tidak terlalu jauh perbedaannya, untuk klasifikasi 2 kelas sel Pap Smear (Normal dan Abnormal) akurasi tertinggi di hasilkan dari metode k-NN (*default k=3*) sebesar 95,05%, sedangkan untuk hasil akurasi dari 7 kelas (*Normal Superficial, Normal Intermediate, Normal Columnar, Mild (Light) Dysplasia, Moderate Dysplasia, Severe Dysplasia, dan Carcinoma In Situ*) dari metode *Decision tree* adalah 64,24% lebih baik di bandingkan dengan metode *Naïve Bayes* dan k-NN. Di bawah ini adalah Gambar 3.1, 3.2 dan 3.3 yang menjelaskan grafik dari hasil eksperimen.



Gambar 3.1 Perbandingan Metode Klasifikasi Normal dan Abnormal



Gambar 3.2 Perbandingan Metode Klasifikasi 7 Kelas



Gambar 3.3 Grafik Perbandingan Hasil Klasifikasi 2 kelas dan 7 kelas

4. Kesimpulan

Berdasarkan dari hasil eksperimen yang telah dilakukan maka dapat di kesimpulan sebagai berikut:

Hasil dari klasifikasi untuk 2 kelas (Normal dan Abnormal) akurasi tertinggi di hasilkan dari metode klasifikasi k-NN sebesar 95,05%; Hasil dari klasifikasi untuk 7 kelas (*Normal Superficial, Normal Intermediate, Normal Columnar, Mild (Light) Dysplasia, Moderate Dysplasia, Severe Dysplasia, dan Carcinoma In Situ*) adalah 64,24% yang di

hasilkan dari metode *Decision Tree*; Hasil eksperimen menunjukkan bahwa kombinasi metode *Particle Swarm Optimization* untuk seleksi fitur dan Teknik Bagging untuk *imbalance Class* terbukti efektif untuk klasifikasi Sel Pap Smear di bandingkan dengan metode optimasi pada penelitian lain. Pada penelitian selanjutnya klasifikasi sel Pap Smear Normal dan Abnormal bisa dilanjutkan ke tahapan implementasi *rule* dari algoritma k-NN yang menghasilkan akurasi baik ke dalam program sehingga dapat memudahkan ahli patologi dalam proses klasifikasi sel Pap Smear (Normal dan Abnormal), sedangkan untuk klasifikasi 7 kelas masih harus mencari metode terbaik untuk klasifikasi Karena akurasi yang di hasilkan pada penelitian ini masih di bawah 80%.

Referensi

- 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.
- Arifin, T. (2014). Implementasi Metode K-Nearest Neighbor Untuk Klasifikasi Citra Sel Pap Smear Menggunakan Analisis Tekstur Nukleus. *Jurnal Informatika*, 2, 287-295.
- Arifin, T. (2014). Metode Data Mining Untuk Klasifikasi Data Sel Nukleus Dan Sel Radang Berdasarkan Analisa Tekstur. *Jurnal Informatika*, 2, 425-433.
- Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques Second Edition*. San Francisco: Diane Cerra.
- Hidayatulloh, T., Herliana, A., & Arifin, T. (2016). Klasifikasi Sel Tunggal Pap Smear Berdasarkan Analisis Fitur Berbasis Naïve Bayes Classifier Dan Particle Swarm Optimization. *SWABUMI*, 4, 186-193.
- Jantzen, J., Norup, G. J., Dounias, & Bjerregaard, B. (2005). Papsmear Benchmark Data For Pattern Classification. *Technical University of Denmark*, 1-20.
- KEMENKES. (2015). *InfoDatin, Pusat Data dan Informasi Kementerian Kesehatan*.
Kementrian Kesehatan Republik Indonesia. (2015). Retrieved July 30, 2016, from www.depkes.go.id: http://www.depkes.go.id/resources/download/pusdatin/infodatin/infodatin-kanker.pdf
- Kusrini, & Luthfi. (2009). *Algoritma Data Mining*. Andi: Yogyakarta.
- Laradji, I. H., Alshayeb, M., & Ghouti, L. (2015). Software Defect Prediction Using Ensemble Learning on Selected Features. *Information and Software Technology*, 388-402.
- Liu, X. Y., & Zhou, Z. H. (2013). Ensemble Methods for Class Imbalance Learning. *Imbalanced Learning: Foundations, Algorithms, and Applications*, First Edition 61-82.
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook Second Edition*. New York: Springer.
- Martin, E. (2013, July 25). *Pap-Smear Classification*. Retrieved from <http://labs.fme.aegean.gr/decision/downloads/>
- Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi.
- Riskesdas. (2013). *Data riset kesehatan*.
- Rodiansyah, & Winarko. (2013). Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive bayesian Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*.
- Septadina, I. S., Kesuma, H., Handayani, D., Suciati, T., & Liana, P. (2007). Upaya Pencegahan Kanker Serviks Melalui Peningkatan Pengetahuan Kesehatan Reproduksi Wanita Dan Pemeriksaan Metode Iva (Inspeksi Visual Asam Asetat) Di Wilayah Kerja Puskesmas Kenten

Palembang. *Jurnal Pengabdian Sriwijaya* .

- Shih Wei, L., Kuo-Ching, Y., Shih-Chieh, C., & Zne-Jung, L. (2008). Particle Swarm Optimization for Parameter Determination and Feature Selection of Support Vector Machines. *ScienceDirect: Expert System With Applications*, 1817-1824.
- Wahono, R. S., & Suryana, N. (2013). Combining Particle Swarm Optimization based Feature Selection and Bagging Technique for Software Defect Prediction. *International Journal of Software Engineering and Its Applications* , 153-166.
- World Health Organization (WHO)*. (2014). Retrieved Agustus 22, 2016, from Cancer: <http://www.who.int/mediacentre/factsheets/fs297/en/>
- Wu, X. (2008). Top 10 algorithms in data mining. Knowledge Inference System. . *IEEE ICDM Conference*, 1-12.