

OPTIMASI KLASIFIKASI SEL TUNGGAL PAP SMEAR MENGUNAKAN CORRELATION BASED FEATURES SELECTION (CFS) BERBASIS C4.5 DAN NAIVE BAYES

Asti Herliana

Fakultas Teknik, Universitas BSI
asti.ala@bsi.ac.id

Abstract

Cervical cancer was the most dangerous disease and generally attacks women. Early detection through Pap Smear method was one way to prevent the disease to grow in womans cervival canal. Based on the result of Pap Smear method, the single cell of data that known as herlev data is available. This data, then used as a reference by the experts to find the best level classification from each class of cervical cancer. The decision tree C4.5 and Naïve Bayes have proven to give the best result on 280 data trial of herlev with support by Correlation based Features Selection (CFS) optimization method. The issues raised in the present study was does CFS optimization method that combined with the classification method of C4.5 and Naïve Bayes can provide increased the accuracy results when it faced the 917 data of herlev. The results of this study show that CFS method that combined either with C4.5 methods and naïve bayes classification accuracy was decrease when compared without using CFS method. In terms of showing that CFS can not provide the best result when if confronted with big data.

Keywords : *optimization, classification, single cell of Pap Smear, Correlation based Features Selection, C4.5, Naïve Bayes*

1. Pendahuluan

Penelitian mengenai kanker serviks berawal dari ditemukannya sebuah metode untuk melakukan pemeriksaan sitologis terhadap leher Rahim yang ditemukan oleh seorang dokter asal Yunani yakni George N. Papanicolaou. Metode pemeriksaan itu sendiri dikenal dengan metode Papanicolaou Smear atau yang kini lebih sering diketahui oleh masyarakat luas dengan metode Pap smear. Berdasarkan metode Pap smear ini, kemudian pada tahun 2003, Erik Martin berhasil membuat sebuah penemuan berupa pembagian kelas sel tunggal Pap Smear menjadi 7 kelas yang terbagi menjadi 3 kelas yang termasuk kelas normal (*Normal Superficial, Normal Intermediate, dan Normal Columnar*) dan 4 kelas yang termasuk kedalam kelas abnormal (*Mild (Light) Dysplasia, Moderate Dysplasia, Severe Dysplasia, dan Carcinoma In Situ*) (Martin, 2003)

Berkembang dari penelitian yang dilakukan oleh Martin, kemudian pada tahun

Dounias dan Beth Bjerregaard melakukan penelitian yang menghasilkan data yang dikenal dengan nama data herlev. Data herlev ini menyajikan hasil segmentasi sel tunggal Pap Smear berupa 917 citra yang terbagi kedalam 7 kelas dengan 20 Fitur seperti *Kerne_A, Cyto_A, KC, Kerne_Ycol, Cyto_Ycol, KerneShort, KerneLong, KerneElong, KerneRund, CytoShort, CytoLong, CytoElong, CytoRund, KernePeri, CytoPeri, KernePos, KerneMax, KerneMin, CytoMax* dan *CytoMin* (Jantzen, Norup, Dounias, & Bjerregaard, 2005). Dengan terciptanya data herlev, banyak peneliti yang melakukan pengembangan dengan meningkatkan hasil akurasi data hasil klasifikasi terhadap masing-masing kelas.

Melihat dari banyaknya fitur yang ada pada data set sel tunggal Pap Smear, dapat dikatakan bahwa data herlev ini merupakan data besar dengan kompleksitas yang cukup besar. Untuk menjawab tantangan akurasi klasifikasi dari data set ini, beberapa metode telah digunakan oleh para ahli dibidang

C4.5, ANFIS, J48, SVM dan Bayesian Classifier. Untuk melakukan peningkatan hasil klasifikasi terhadap 2 kelas (normal dan abnormal) dan 7 kelas (*Normal Superficial, Normal Intermediate, Normal Columnar, Mild (Light) Dysplasia, Moderate Dysplasia, Severe Dysplasia, dan Carcinoma In Situ*) dari sel tunggal Pap Smear, terdapat beberapa pendekatan metode optimasi yang juga telah dilakukan. Diantara beberapa metode optimasi yang telah digunakan oleh para ahli, metode *Correlation based Features Selection* (CFS), menunjukkan hasil yang lebih baik berdasarkan penelitian yang dilakukan oleh Izzati Muhummah, Komariyuli Anwariyah dan Indrayanti (Muhummah, Anwariyah, & Indrayanti, 2012). Penelitian selanjutnya dilakukan oleh Herliana dan Riana dimana pada penelitian ini digunakan optimasi CFS berbasis *Decision Tree* J48 terhadap analisis fitur dan analisis tekstur sel tunggal Pap Smear dengan jumlah data sebanyak 280. Hasil dari penelitian ini menunjukkan peningkatan hasil akurasi klasifikasi menjadi sebesar 90% untuk 2 kelas dan sebesar 67,87% untuk 7 kelas (Herliana & Riana, 2013). Selain itu, pada tahun 2014 Arifin dapat membuktikan bahwa dengan menggunakan metode algoritma C4.5 yang dioptimasi menggunakan metode CFS pada data citra dari data set Herlev dengan mengambil 280 citra, hasil akurasi klasifikasi yang didapatkan dapat meningkat hingga mencapai akurasi 96,43% (Arifin, 2014).

Untuk mengetahui apakah metode optimasi CFS dapat memberikan hasil yang lebih maksimal pada Analisis Fitur sel tunggal Pap Smear dengan data 917 sama baiknya ketika dilakukan pengujian terhadap data 280, maka pada penelitian kali ini dilakukan perbandingan hasil dari dua metode klasifikasi terpilih yakni Naïve Bayes dan C4.5. Kedua metode ini akan diujikan terhadap data 917 dari sel tunggal Pap Smear untuk didapatkan hasil optimal yang dapat dicapai dari kedua metode jika diberikan metode optimasi CFS.

Landasan Teori

a. Naïve Bayes

Metode Naïve Bayes merupakan metode yang lebih handal jika dibandingkan dengan *Support Vector Machine* (SVM) dalam hal pengolahan data yang besar dengan fitur yang beragam. Kelebihan dari metode Naïve Bayes adalah metode ini merupakan algoritma yang sederhana namun memiliki akurasi yang tinggi. (Rodiyansyah & Winarko,

2012). Sedangkan kekurangan dari metode ini adalah adanya kesulitan dalam hal kecocokan terhadap suatu aturan yang kurang baik. Dengan kata lain, jika terdapat suatu aturan yang diterapkan pada sebuah algoritma seperti penggunaan data angka riil, maka akan sulit untuk menentukan priornya. Secara garis besar, algoritma dari Naïve Bayes digambarkan sebagai berikut:

$$p(I=i_j|C=c_i) = \frac{p(I=i_j) p(C=c_i|I=i_j)}{p(C=c_i)} \dots\dots (1)$$

Keterangan :

$p(I=i_j|C=c_i)$: peluang interval i ke- j untuk kelas c_i

$p(C=c_i|I=i_j)$: peluang kelas c_i pada interval i ke- j

$p(I=i_j)$: peluang sebuah interval ke- j pada semua interval yang terbentuk

$p(C=c_i)$: peluang sebuah kelas ke- i untuk semua kelas yang ada di dataset

b. C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomizer Three*) yang digunakan untuk membuat sebuah pohon keputusan (Larose, 2005). Sama halnya seperti algoritma *decision tree* pendahulunya, C4.5 memiliki *root node, internal node* dan *leaf node*. Menurut Kusriani dan Taufiq, dalam pemilihan *node root*, maka akan dipilih atribut dengan nilai *gain* tertinggi (Larose, 2005). Nilai *gain* dan *entropy* yang terbentuk dalam sebuah *decision tree* didapatkan melalui persamaan sebagai berikut:

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \cdot \text{Entropy}(S_i) \dots\dots(2)$$

Keterangan :

S : Himpunan Kasus

A : Atribut

n : Jumlah Partisi Atribut A

$|S_i|$: Jumlah Kasus pada Partisi ke i

$|S|$: Jumlah Kasus dalam S

Untuk mengetahui Entropy maka digunakan rumus sebagai berikut:

$$\text{Entropy}(S) = - \sum_{i=1}^n p_i \cdot \log_2 p_i \dots\dots(3)$$

Keterangan :

S : Himpunan Kasus

A : Fitur

n : Jumlah Partisi S
 p_i : Proporsi dari S_i terhadap S

c. *Correlation based Features Selection* (CFS)

CFS merupakan salah satu metode optimasi seleksi atribut. Metode yang digunakan dalam metode ini adalah dengan cara menghitung dan membandingkan tingkat korelasi antara masing-masing variabel terhadap variabel kelasnya dan antara atribut itu sendiri (Ompusunggu, 2008). Algoritma dasar pembentukan *decision tree* pada algoritma C4.5 dapat dinyatakan sebagai berikut (Han & Kamber, 2001):

Input : *Training samples, Attribute*

Output: *Decision Tree*

Generate *decision_tree* (*Training samples, Attribute*) // fungsi *decision tree*

Method:

- (1) Buat sebuah node N;
- (2) **If** sekumpulan data dari sampel terdiri dari kelas yang sama dengan C, **then**
- (3) **Return** N sebagai *leaf node* dengan label kelas C;
- (4) **If** *atribut_list* kosong **then**
- (5) **Return** N sebagai *leaf node* dengan label kelas mayoritas dalam sampel; // mayoritas voting
- (6) **Else**
- (7) Terapkan metode *seleksi atribut*, untuk mendapatkan *splitting_kriteria* terbaik
- (8) Berikan label node N dengan *splitting_kriteria*;
- (9) **For each** hasil a_1 dari *splitting_kriteria* // partisi dari sampel
- (10) Tumbuhkan cabang dari node N untuk kondisi *splitting_kriteria* = a_1 ;
- (11) Biarkan s_1 menjadi set dari sampel pada sampel **for** dimana *splitting_kriteria* = a_1 ; sebuah partisi
- (12) **If** s_1 kosong **then**
- (13) Tambahkan sebuah *leaf* dengan label mayoritas kelas di sampel Untuk node N;
Else ambil node kembali dengan **Generate_decision_tree** (s_1 .
Atribut_list) untuk node N;
Endfor

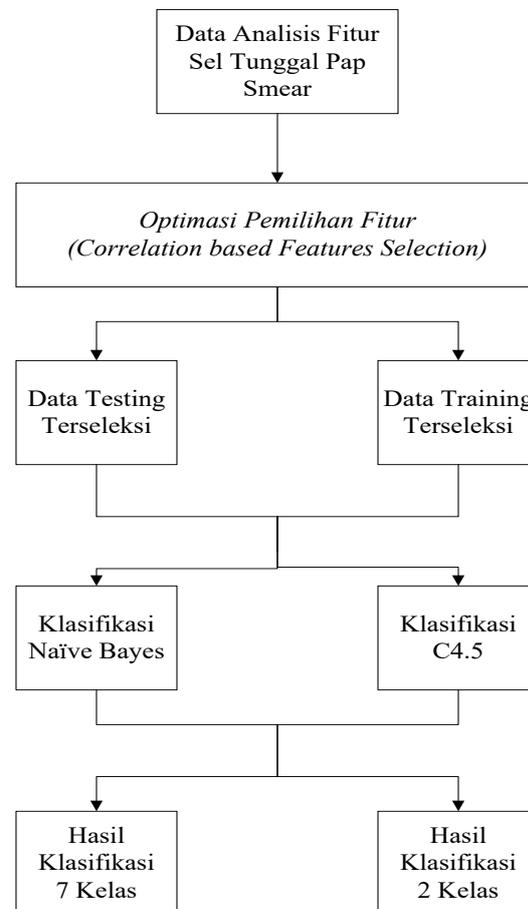
2. Metode Penelitian

Material yang digunakan pada penelitian kali ini adalah data numeric dari data set herlev yang terdiri dari 917 data yang masing-masing data memiliki 20 fitur. Data set

ini sebelumnya dibagi menjadi data testing dan data training.

Untuk proses klasifikasi pada penelitian ini dilakukan terhadap 7 kelas sel tunggal Pap Smear (*Normal Superficial, Normal Intermediate, Normal Columnar, Mild (Light) Dysplasia, Severe Dysplasia, Moderate Dysplasia* dan *Carcinoma In Situ*). dan 2 kelas sel tunggal *Pap Smear* yang terdiri dari kelas Normal dan kelas Abnormal.

Untuk mempermudah penggambaran dari proses penelitian kali ini, digambarkan melalui gambaran metode penelitian sebagai berikut :



Gambar 1. Metodologi Penelitian

Berdasarkan gambaran metodologi penelitian diatas, dapat terlihat bahwa data herlev pertama kali akan dioptimasi dengan maksud untuk mendapatkan fitur terbaik dari 20 fitur yang ada. Selanjutnya, data fitur terseleksi akan dibagi menjadi data training dan data testing. Setelah didapatkan data testing dan data training, kemudian

selanjutnya akan dilakukan klasifikasi dengan menggunakan metode C4.5 dan metode Naïve Bayes untuk mendapatkan hasil klasifikasi terhadap sel tunggal Pap Smear tersebut.

Berdasarkan hasil dari seleksi fitur dengan menggunakan CFS, maka didapatkan

fitur terbaik untuk penelitian kali ini sebanyak 5 fitur yakni *Kerne_A*, *Cyto_A*, *K/C*, *KerneLong*, *KerneMax* yang digambarkan sebagai berikut :

```

=== Attribute Selection on all input data ===

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 162
  Merit of best subset found: 0.529

Attribute Subset Evaluator (supervised, Class (nominal): 21 Class):
  CFS Subset Evaluator
  Including locally predictive attributes

Selected attributes: 1,2,3,7,17 : 5
  Kerne_A
  Cyto_A
  K/C
  KerneLong
  KerneMax

```

Gambar 2 Hasil Seleksi Fitur Menggunakan CFS

Setelah didapatkan data fitur terbaik, proses selanjutnya adalah mencari hasil klasifikasi.

Berikut ini digambarkan hasil *running* proses klasifikasi terhadap 2 kelas

```

Time taken to build model: 0.08 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      845      92.1483 %
Incorrectly Classified Instances    72       7.8517 %
Kappa statistic                    0.7946
Mean absolute error                 0.0875
Root mean squared error             0.2684
Relative absolute error             22.501 %
Root relative squared error         60.8999 %
Total Number of Instances          917

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.826   0.044   0.870   0.826   0.847   0.795   0.901   0.837   normal
0.956   0.174   0.939   0.956   0.947   0.795   0.901   0.920   abnormal
Weighted Avg.   0.921   0.139   0.921   0.921   0.921   0.795   0.901   0.898

=== Confusion Matrix ===

  a  b  <-- classified as
200 42 |  a = normal
 30 645 |  b = abnormal

```

Gambar 3 Hasil Klasifikasi C4.5 Tanpa Optimasi CFS

```

Time taken to build model: 0.09 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      844          92.0393 %
Incorrectly Classified Instances    73           7.9607 %
Kappa statistic                    0.7887
Mean absolute error                0.1196
Root mean squared error            0.2643
Relative absolute error            30.7507 %
Root relative squared error        59.9763 %
Total Number of Instances          917

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.802   0.037   0.886     0.802   0.842     0.790   0.905    0.871    normal
                0.963   0.198   0.931     0.963   0.947     0.790   0.905    0.933    abnormal
Weighted Avg.   0.920   0.156   0.919     0.920   0.919     0.790   0.905    0.917

=== Confusion Matrix ===

  a  b  <-- classified as
194 48 | a = normal
 25 650 | b = abnormal

```

Gambar 4 Hasil Klasifikasi C4.5 Dengan Optimasi CFS

```

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      841          91.7121 %
Incorrectly Classified Instances    76           8.2879 %
Kappa statistic                    0.7715
Mean absolute error                0.0881
Root mean squared error            0.2759
Relative absolute error            22.6515 %
Root relative squared error        62.5951 %
Total Number of Instances          917

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.740   0.019   0.932     0.740   0.825     0.780   0.956    0.915    normal
                0.981   0.260   0.913     0.981   0.946     0.780   0.956    0.981    abnormal
Weighted Avg.   0.917   0.197   0.918     0.917   0.914     0.780   0.956    0.963

=== Confusion Matrix ===

  a  b  <-- classified as
179 63 | a = normal
 13 662 | b = abnormal

```

Gambar 5 Hasil Klasifikasi Naïve Bayes Tanpa Optimasi CFS

```

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      845          92.1483 %
Incorrectly Classified Instances    72           7.8517 %
Kappa statistic                    0.7963
Mean absolute error                 0.0941
Root mean squared error             0.2553
Relative absolute error             24.2006 %
Root relative squared error         57.9179 %
Total Number of Instances          917

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.839   0.049   0.860     0.839   0.849     0.796   0.960    0.929   normal
                0.951   0.161   0.943     0.951   0.947     0.796   0.960    0.981   abnormal
Weighted Avg.   0.921   0.132   0.921     0.921   0.921     0.796   0.960    0.967

=== Confusion Matrix ===

  a  b  <-- classified as
203 39 | a = normal
 33 642 | b = abnormal

```

Gambar 5 Hasil Klasifikasi Naïve Bayes Dengan Optimasi CFS

3. Hasil Dan Pembahasan

Setelah didapatkan fitur terbaik dari data set sel tunggal Pap Smear, selanjutnya data dibagi menjadi data testing dan data training. Setelah dibagi, maka data set kemudian diujikan terhadap algoritma Naïve Bayes dan C4.5. Nilai prosentase akurasi didapatkan dari rumus sebagai berikut:

$$\text{Akurasi} = \frac{\text{Prediksi Benar}}{\text{Total Data Prediksi}} \times 100\%$$

Berdasarkan hasil olah data didapatkan hasil sebagai berikut :

Tabel I Hasil Akurasi Klasifikasi 2 Kelas (Naïve Bayes)

Normal dan Abnormal			
Algoritma	Training	Testing	Total Akurasi
NBC	90	10	91,3%
NBC	80	20	92,9%
NBC	70	30	89,6%
NBC	60	40	90,79%
NBC	50	50	97,4%

Tabel II Hasil Akurasi Klasifikasi 7 Kelas (Naive Bayes)

Normal dan Abnormal			
Algoritma	Training	Testing	Total Akurasi
NBC + CFS	90	10	90,65%
NBC + CFS	80	20	91,26%
NBC + CFS	70	30	89,75%
NBC + CFS	60	40	90,87%
NBC + CFS	50	50	88,91%

Tabel III Hasil Akurasi Klasifikasi 2 Kelas (Naïve Bayes + CFS)

Normal dan Abnormal			
Algoritma	Trainin g	Testin g	Total Akurasi
NBC + CFS	90	10	90,65%
NBC + CFS	80	20	91,26%

NBC + CFS	70	30	89,75%
NBC + CFS	60	40	90,87%
NBC + CFS	50	50	88,91%

Tabel IV Hasil Akurasi Klasifikasi 7 Kelas (Naïve Bayes + CFS)

KELAS 1-7			
Algoritma	Trainin g	Testin g	Total Akurasi
NBC + CFS	90	10	53,80%
NBC + CFS	80	20	58,47%
NBC + CFS	70	30	53,05%
NBC + CFS	60	40	55,10%
NBC + CFS	50	50	51,86%

Tabel V Hasil Akurasi Klasifikasi 2 Kelas (C4.5)

Normal dan Abnormal			
Algoritma	Trainin g	Testin g	Total Akurasi
C4.5	90	10	95,65%
C4.5	80	20	94,37%
C4.5	70	30	89,67%
C4.5	60	40	89,86%
C4.5	50	50	89,35%

Tabel VI Hasil Akurasi Klasifikasi 7 kelas (C4.5)

KELAS 1-7			
Algoritma	Training	Test	Total Akurasi
C4.5	90	10	52,28%
C4.5	80	20	50,71%
C4.5	70	30	49,09%
C4.5	60	40	51,39%
C4.5	50	50	50,35%

Tabel VII Hasil Akurasi Klasifikasi 2 Kelas (C4.5 + CFS)

Normal dan Abnormal			
Algoritma	Trainin g	Testin g	Total Akurasi
C4.5+ CFS	90	10	92,39%
C4.5+ CFS	80	20	94,97%
C4.5+ CFS	70	30	89,85%
C4.5+ CFS	60	40	91,61%
C4.5+ CFS	50	50	90%

Tabel VIII Hasil Akurasi Klasifikasi 7 Kelas (C4.5 + CFS)

KELAS 1-7			
Algoritma	Trainin g	Testing	Total Akurasi
C4.5+ CFS	90	10	53,70%
C4.5+ CFS	80	20	58,09%
C4.5+ CFS	70	30	52,58%
C4.5+ CFS	60	40	55,40%
C4.5+ CFS	50	50	50,41%

Berdasarkan hasil perhitungan klasifikasi yang disampaikan pada tabel I – tabel VIII, didapatkan bahwa untuk fitur yang beragam dan data yang besar, metode CFS tidak dapat memberikan hasil yang lebih baik. Selain itu, metode naïve bayes terbukti lebih unggul dari pada metode C4.5 dalam menghasilkan hasil akurasi terhadap data set sel tunggal *Pap Smear* yang memiliki 917 data dengan 20 fitur baik untuk 2 kelas maupun 7 kelas.

Hasil dari penelitian ini dapat dijadikan acuan bagi para peneliti selanjutnya dalam menggunakan metode yang tepat terhadap data set lain untuk dilakukan optimasi dan klasifikasi. Selain itu, penelitian ini juga dapat dijadikan salah satu referensi bagi para ahli dalam bidang bioinformatika dalam menentukan dan meningkatkan kualitas dari data yang telah ada.

4. Kesimpulan

Dalam penelitian kali ini dilakukan perbandingan metode klasifikasi Naïve Bayes dengan C4.5 yang dioptimasi menggunakan CFS. Dari hasil pembahasan didapatkan beberapa kesimpulan sebagai berikut:

1. Penggunaan data keseluruhan dari sel tunggal Pap Smear yakni sebesar 917 data menunjukkan tingkat akurasi yang menurun jika menggunakan metode optimasi CFS. Namun jika menggunakan data sebesar 280 seperti data yang digunakan pada penelitian Herliana dan Riana, maka hasil akurasi cenderung meningkat.
2. Metode *Correlation-based Features Selection* (CFS) tidak memberikan performa pendukung terbaik jika diterapkan kepada data yang besar seperti data herlev ini.
3. Hasil prediksi terhadap 2 kelas dan 7 kelas mengalami penurunan prediksi jika dibandingkan dengan hasil klasifikasi tanpa menggunakan metode optimasi CFS.
4. Algoritma C4.5 menunjukkan hasil lebih baik untuk klasifikasi terhadap 2 kelas, sedangkan untuk hasil klasifikasi terhadap 7 kelas Naïve Bayes menunjukkan keunggulannya terhadap C4.5.

Saran

1. Bagi penelitian selanjutnya, jika akan menggunakan 917 data, disarankan menggunakan metode optimasi *Particle Swarm Optimization* (PSO) dan metode Naïve Bayes sebagai metode klasifikasinya.
2. Fitur dari sel tunggal *Pap Smear* dapat dikembangkan tidak hanya mengacu pada 20 fitur dari data herlev saja.

Referensi

- Arifin, T. (2014). Klasifikasi Inti Sel Pap Smear Berdasarkan Analisis Tekstur Menggunakan Correlation Based Features Selection Berbasis Algoritma C4.5. *Jurnal Informatika*. Vol. II No. 2, 123-129.
- Han, J., & Kamber, M. (2001). *Data Mining Concepts and Techniques*. The Morgan Kaufmann Series.
- Herliana, A., & Riana, D. (2013, November 3). *Klasifikasi Sel Tunggal Pap Smear Berdasarkan Analisis Fitur Dan Analisis Tekstur Terseleksi Menggunakan Correlation-Based Features Selection*. Jakarta: Stmik Nusa Mandiri. Retrieved From [Www.Academia.Edu: Https://Www.Academia.Edu/11538482/Klasifikasi_Sel_Tunggal_Pap_Sme_ar_Berdasarkan_Analisis_Fitur_Dan_Analisis_Tekstur_Terseleksi_Menggunakan_Correlation-Based_Features_Selection](https://www.academia.edu/11538482/Klasifikasi_Sel_Tunggal_Pap_Sme_ar_Berdasarkan_Analisis_Fitur_Dan_Analisis_Tekstur_Terseleksi_Menggunakan_Correlation-Based_Features_Selection)
- Jantzen, J., Norup, J., Dounias, G., & Bjerregaard, B. (2005). *Pap-Smear Benchmark Data For Pattern Classification*. Harlev: Technical University of Denmark.
- Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: Jhon Wiley & Sons Inc.
- Martin, Erik. (2003). *Pap-Smear Classification*. Juni 19, 2013, from Technical University of Denmark: <http://labs.fme.aegean.gr/decision/downloads/>
- Muhummah, I., Anwariyah, K., & Indrayanti. (2012). Extraction and Selection Features of Cervical Cell Types in Papsmears Digital Images. *International Conference on Woman's Health in Science & Engineering (WISE-Health) 2012*. Bandung: Institute Technology Bandung.
- Ompusunggu, H. S. (2008). *Implementasi Seleksi Variabel Menggunakan Metode Correlation-Based Feature Selection Dalam Task Classification*. Bandung: Institut Teknologi Telkom.
- Rodiyansyah, S. F., & Winarko, E. (2012). Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. *IJCCS Vol.6 No.1 ISSN: 1978-1520*, 91-100.