

**Pengaruh *Forward Selection* Pada Klasifikasi Menggunakan Algoritma *Naïve Bayes***

Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Studi Strata-1 Pada  
Jurusan Teknik Informatika Reguler Fakultas Ilmu Komputer



Oleh :

**MISBAHUDIN MUNADI**

**09021281520118**

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2019**

**Pengaruh *Forward Selection* Pada Klasifikasi Menggunakan Algoritma *Naïve Bayes***

Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Studi Strata-1 Pada  
Jurusan Teknik Informatika Reguler Fakultas Ilmu Komputer



Oleh :

Misbahudin Munadi  
09021281520118

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2019**

**LEMBAR PENGESAHAN TUGAS AKHIR**


**PENGARUH FORWARD SELECTION PADA KLASIFIKASI  
MENGUNAKAN ALGORITMA NAIVE BAYES**

Oleh :

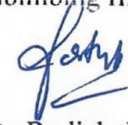
**MISBAHUDIN MUNADI**

NIM : 09021281520118

Pembimbing I,

  
Novi Yusliani, M.T  
NIP. 198211082012122001

Indralaya, 23 Desember 2019  
Pembimbing II,

  
Desty Rodiah, M.T  
NIK. 1671016112890005

Mengetahui,  
Ketua Jurusan,

  
Rifkie Primartha, S.T., M.T.  
NIP. 197706012009121004

## TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari **Jumat** tanggal **18 Desember 2019** telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Misbahudin Munadi

NIM : 09021281520118

Judul : Pengaruh *Forward Selection* Pada Klasifikasi Menggunakan Algoritma *Naïve Bayes*

1. Pembimbing I  
Novi Yusliani, M.T. ....  
NIP. 198211082012122001
2. Pembimbing II  
Desty Rodiah, M.T. ....  
NIK. 1671016112890005
3. Penguji I  
Yunita, M.Cs.....  
NIP. 198306062015042002
4. Penguji II  
Dany Matthew Saputra, M.Sc.....  
NIP. 198505102015041002

Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Primartha S.T. M.T.

NIP 197706012009121004

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Misbahudin Munadi

NIM : 09021281520118

Program Studi : Teknik Informatika

Judul Skripsi : Pengaruh *Forward Selection* pada Klasifikasi Menggunakan Metode *Naive Bayes*

Hasil Pengecekan Software *iThenticate/Turnitin* : 8%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Indralaya, 23 Desember 2019



Misbahudin Munadi

NIM. 09021281520118

*Motto:*

Rasulullah *Shallallahu 'alaihi Wasallam* yang bersabda:

*“Barangsiapa yang (menjadikan) dunia tujuannya maka Allah akan mencerai-beraikan urusannya dan menjadikan kemiskinan/tidak pernah merasa cukup (selalu ada) di hadapannya, padahal dia tidak akan mendapatkan (harta benda) duniawi melebihi dari apa yang Allah tetapkan baginya. Dan barangsiapa yang (menjadikan) akhirat niatnya maka Allah akan menghimpunkan urusannya, menjadikan kekayaan/selalu merasa cukup (ada) dalam hatinya, dan (harta benda) duniawi datang kepadanya dalam keadaan rendah (tidak bernilai dihadapannya)”*

*Kupersembahkan karya tulis ini kepada:*

- *Orang tuaku tersayang*
- *Keluarga besarku*
- *Sahabat dan teman seperjuanganku*
- *Fakultas Ilmu Komputer*
- *Universitas Sriwijaya*

THE EFFECT OF FORWARD SELECTION ON  
CLASSIFICATION USING NAIVE BAYES

By:

Misbahudin Munadi  
09021281520118

ABSTRACT

Naive Bayes is a classification method that has an independence assumption for classification, but in fact, independent assumptions on attributes almost never happen. Therefore, this study will examine the effect of forward selection, which is a forward selection method that depends on dependency of each attribute. Then the attribute model will be evaluation and compared with without the selection feature. Based on the evaluation results conducted produces effects using harmonic mean values of F Measure and accuracy on the naive bayes method, and the naive bayes method with forward selection on the Pima Indians Diabetes dataset of 0.7391, and 0.7538. In the Breast Cancer dataset of 0.7120, and 0.7120. And the SPECTF Heart dataset is 0.5453, and 0.5962. Forward selection with the collaboration between the dependent attributes to produce a simpler attribute model and shows that forward selection has the effect of optimizing the naive bayes classification method, naive bayes with forward selection can work with simpler attributes.

**Keywords:** *Classification, Naive Bayes, Forward Selection, Pearson Correlation.*

Pembimbing I,

Novi Yuslianti, M.T  
NIP. 198211082012122001

Indralaya, 23 Desember 2019  
Pembimbing II,

Desty Rodiah, M.T  
NIK. 1671016112890005

Mengetahui,  
Ketua Jurusan,



Rifkie Primartha, S.T., M.T.  
NIP. 197706012009121004

PENGARUH FORWARD SELECTION PADA KLASIFIKASI  
MENGUNAKAN ALGORITMA NAIVE BAYES

Oleh:

Misbahudin Munadi  
09021281520118

**ABSTRAK**

*Naive Bayes* merupakan salah satu metode *classification* yang memiliki asumsi independensi untuk pengklasifikasiannya, namun kenyataannya asumsi independen pada atribut hampir tidak pernah terjadi. Oleh karena itu, pada penelitian ini akan menguji pengaruh *forward selection* yaitu metode fitur seleksi yang bergantung pada dependensi antar setiap atribut. Dimana nantinya setiap model atribut yang terbentuk akan di uji dan dibandingkan dengan tanpa fitur seleksi. Berdasarkan penelitian yang dilakukan menghasilkan pengaruh dengan menggunakan nilai rerata harmonik *F Measure* dan *accuracy* pada metode *naive bayes*, dan metode *naive bayes* dengan *forward selection* pada dataset Pima Indians Diabetes sebesar 0,7391, dan 0,7538. Pada dataset Breast Cancer sebesar 0,7120, dan 0,7120. Dan pada dataset SPECTF Heart sebesar 0,5453, dan 0,5962. *Forward selection* yaitu dengan kerjasama antar atribut dependen menghasilkan model atribut yang lebih sederhana dan menunjukkan bahwa *forward selection* berpengaruh terhadap pengoptimalan metode klasifikasi *naive bayes*, dimana *naive bayes* dengan *forward selection* dapat bekerja dengan kompleksitas atribut yang lebih sederhana.

**Keywords:** *Classification, Naive Bayes, Forward Selection, Pearson Correlation.*

Pembimbing I,

Novi Yuliani, M.T  
NIP. 1982111082012122001

Indralaya, 23 Desember 2019  
Pembimbing II,

Desty Rodiah, M.T  
NIK. 1671016112890005

Mengetahui,  
Ketua Jurusan,

Rifkie Primartha, S.T., M.T.  
NIP. 197706012009121004



## KATA PENGANTAR



Puji syukur kepada Allah SWT atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Allah Subhanahu wa Ta'ala yang telah melimpahkan kepada hamba keimanan, kesehatan, kecerdasan, kemudahan, dan kelancaran sehingga hamba dapat menyelesaikan tugas-tugas hamba sebagai seorang mahasiswa.
2. Nabi Muhammad Shallallahu 'Alaihi wa Sallam yang telah memberikan jalan yang terang dan memberikan ilmu yang bermanfaat kepada umatnya atas seizin Allah.
3. Bapak Jaidan Jauhari, S.Pd., M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Rifkie Primartha, M.T selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya serta Ibu Hardini Novianti, M.T. selaku Sekretaris Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Novi Yusliani, M.T dan Buk Desty Rodiah, M.T selaku pembimbing Tugas Akhir saya yang telah mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
6. Ibu Yunita, M.Cs selaku dosen penguji I dan Bapak Danny Matthew Saputra, M.Sc selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
7. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Orangtua saya, Wasito Sutarno dan Romlah yang telah memberikan dorongan dan dukungan baik secara moril maupun materil, serta kepada ayuk saya saudari Hijir Andila Nurohmi beserta suaminya Ahmad Fikky serta anaknya Fidi Khoirul Putra dan saudari Atika Rahmawati.

9. Kak Ricy dan Mbak Winda serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
10. Bang Ardi Wasila Chandra, Wak Tirta, M. Agung Pribadi, Kerel Khalif Afif, Albar, Genda Ananta, Gian Agnar, Rizky Blend, Bang Rasyid Hilmi, Rayyan, Arfah, Dian Ayu, Zakia, Destrilia, Rani, Divinda dan teman-teman IF REG B 2015, Videografi Unsri. Serta seluruh teman-teman Teknik Informatika yang telah berbagi cerita kehidupan bersama pada masa perkuliahan ini.
11. Video pembelajaran online pemrograman yang telah membantu penulis dalam menimba ilmu dalam pemrograman.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya.

Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, 01 Desember 2019

Misbahudin Munadi

## DAFTAR ISI

|                                    |      |
|------------------------------------|------|
| LEMBAR PENGESAHAN TUGAS AKHIR..... | ii   |
| ABSTRACT .....                     | vi   |
| ABSTRAK .....                      | vii  |
| KATA PENGANTAR .....               | viii |
| DAFTAR ISI .....                   | x    |
| DAFTAR TABEL.....                  | xiv  |
| DAFTAR GAMBAR.....                 | xvii |
| BAB I.....                         | I-1  |
| PENDAHULUAN.....                   | 1-1  |
| Pendahuluan.....                   | I-1  |
| 1.2 Latar Belakang .....           | I-1  |
| 1.3 Rumusan Masalah .....          | I-3  |
| 1.4 Tujuan Penelitian.....         | I-4  |
| 1.5 Manfaat Penelitian.....        | I-4  |
| 1.6 Batasan Masalah.....           | I-5  |
| 1.7 Sistematika Penulisan .....    | I-5  |
| 1.8 Kesimpulan .....               | I-8  |
| BAB II.....                        | II-1 |
| KAJIAN TEORITIS.....               | II-1 |
| 2.1 Pendahuluan.....               | II-1 |
| 2.2 Landasan Teori.....            | II-1 |
| 2.2.1 Klasifikasi .....            | II-1 |

|                                                                            |              |
|----------------------------------------------------------------------------|--------------|
| 2.2.2 Fitur Seleksi .....                                                  | II-2         |
| 2.2.3 Pearson Correlation .....                                            | II-4         |
| 2.2.4 Forward Selection.....                                               | II-4         |
| 2.2.5 Naïve Bayes Gaussian .....                                           | II-6         |
| 2.2.6 K-Fold Cross Validation .....                                        | II-8         |
| 2.2.7 Confusion Matrix.....                                                | II-9         |
| 2.3 Penelitian Lain Yang Relevan .....                                     | II-11        |
| 2.4 Kesimpulan .....                                                       | II-14        |
| <b>BAB III .....</b>                                                       | <b>III-1</b> |
| <b>METODOLOGI PENELITIAN .....</b>                                         | <b>III-1</b> |
| 1 3.1 Pendahuluan.....                                                     | III-1        |
| 3.2 Metode Pengumpulan Data.....                                           | III-1        |
| 3.3 Tahapan Penelitian .....                                               | III-2        |
| 3.3.1 Menetapkan Kerangka Kerja.....                                       | III-2        |
| 3.3.2 Menetapkan Kriteria Pengujian.....                                   | III-4        |
| 3.3.3 Menetapkan Format Data Pengujian .....                               | III-4        |
| 3.3.4 Menentukan Alat Yang Digunakan Dalam Pelaksanaan Penelitian<br>..... | III-5        |
| 3.3.5 Melakukan Pengujian Penelitian .....                                 | III-6        |
| 3.3.6 Melakukan Analisis Hasil Pengujian dan Kesimpulan.....               | III-6        |
| 3.4 Metode Pengembangan Perangkat Lunak .....                              | III-7        |
| 3.4.1 Rational Unified Process.....                                        | III-7        |
| 3.4.1.1 Fase Insepsi .....                                                 | III-8        |
| 3.4.1.2 Fase Elaborasi.....                                                | III-9        |

|                                                                     |        |
|---------------------------------------------------------------------|--------|
| 3.4.1.3 Fase Konstruksi .....                                       | III-9  |
| 3.4.1.4 Fase Transisi .....                                         | III-10 |
| 3.5 Manajemen Proyek Penelitian .....                               | III-10 |
| BAB IV .....                                                        | IV-1   |
| PENGEMBANGAN PERANGKAT LUNAK .....                                  | IV-1   |
| 4.1.Pendahuluan .....                                               | IV-1   |
| 4.2.Fase Insepsi .....                                              | IV-1   |
| 4.2.1 Analisis Perangkat Lunak .....                                | IV-2   |
| 4.2.1.1 Analisis Kebutuhan Perangkat Lunak .....                    | IV-2   |
| 4.2.1.2 Analisis Data .....                                         | IV-2   |
| 4.2.1.3 Analisis Korelasi Terhadap Variabel y .....                 | IV-3   |
| 4.2.1.4 Analisis Linier Regresi Bertahap .....                      | IV-4   |
| 4.2.1.5 Analisis Naive Bayes .....                                  | IV-5   |
| 4.2.2 Desain Perangkat Lunak .....                                  | IV-6   |
| 4.2.2.1 Kebutuhan Sistem .....                                      | IV-8   |
| 4.2.2.2 Fungsi Memuat Data .....                                    | IV-8   |
| 4.2.2.3 Fungsi Klasifikasi Naive Bayes .....                        | IV-10  |
| 4.2.1.1. Fungsi Klasifikasi Naive Bayes dan Forward Selection ..... | IV-11  |
| 4.3 Fase Elaborasi .....                                            | IV-12  |
| 4.3.1 Kebutuhan Sistem .....                                        | IV-12  |
| 4.3.1.1 Muat Data .....                                             | IV-13  |
| 4.3.1.2 Fitur Model Klasifikasi .....                               | IV-14  |
| 4.3.2 Pemodelan Bisnis .....                                        | IV-15  |
| 4.3.3 Analisis Desain .....                                         | IV-18  |

|                                                                                                                             |            |
|-----------------------------------------------------------------------------------------------------------------------------|------------|
| 4.3 Fase Konstruksi .....                                                                                                   | IV-20      |
| 4.3.4 Pemodelan Bisnis .....                                                                                                | IV-20      |
| 4.3.5 Kebutuhan Sistem .....                                                                                                | IV-22      |
| 4.4 Fase Transisi.....                                                                                                      | IV-23      |
| 4.4.1 Pemodelan Bisnis .....                                                                                                | IV-23      |
| 4.4.2 Kebutuhan.....                                                                                                        | IV-23      |
| 4.4.3 Pengujian .....                                                                                                       | IV-24      |
| 4.4.3.1 Rencana Pengujian .....                                                                                             | IV-24      |
| 4.4.3.2 Pengujian Perangkat Lunak.....                                                                                      | IV-25      |
| 4.5 Kesimpulan .....                                                                                                        | IV-29      |
| <b>BAB V .....</b>                                                                                                          | <b>V-1</b> |
| <b>HASIL DAN ANALISIS PENELITIAN .....</b>                                                                                  | <b>V-1</b> |
| 5.1 Pendahuluan .....                                                                                                       | V-1        |
| 5.2. Data Uji Hasil Penelitian .....                                                                                        | V-1        |
| 5.2.1 Hasil Pengujian Pengklasifikasian pada Teknik Naive Bayes. V-1                                                        |            |
| 5.2.1.1 Hasil Pengujian Klasifikasi Naive Bayes Terhadap Dataset<br>Diabetes pada Suku Pima Indians .....                   | V-1        |
| 5.2.1.2 Hasil Pengujian Klasifikasi Naive Bayes Terhadap Dataset<br>Diabetes pada Suku Pima Indians. ....                   | V-2        |
| 5.2.1.3 Hasil Pengujian Klasifikasi Naive Bayes Terhadap Dataset<br>SPECTF Heart.....                                       | V-3        |
| 5.2.2 Hasil Pengujian Pengklasifikasian pada Teknik Naive Bayes dan<br>Forward Selection.....                               | V-4        |
| 5.2.2.1 Hasil Pengujian Klasifikasi Naive Bayes dan Forward<br>Selection Terhadap Dataset Diabetes pada suku Pima Indians . | V-5        |

|                                                                                                            |      |
|------------------------------------------------------------------------------------------------------------|------|
| 5.2.2.2 Hasil Pengujian Klasifikasi Naive Bayes dan Forward Selection Terhadap Dataset Breast Cancer ..... | V-13 |
| 5.2.2.3 Hasil Pengujian Klasifikasi Naive Bayes dan Forward Selection Terhadap Dataset SPECTF Heart .....  | V-23 |
| 5.2.3 Hasil Pengujian Hasil Analisis Klasifikasi .....                                                     | V-67 |
| 5.2.3.1 Perbandingan Algoritma Klasifikasi pada Dataset Diabetes pada Suku Pima Indians .....              | V-67 |
| 5.2.3.2 Perbandingan Algoritma Klasifikasi pada Dataset Breast Cancer .....                                | V-68 |
| 5.2.3.3 Perbandingan Algoritma Klasifikasi pada Dataset SPECTF Heart .....                                 | V-70 |
| 5.2.4 Analisis Hasil Uji Penelitian .....                                                                  | V-71 |
| 5.3 Kesimpulan .....                                                                                       | V-73 |
| BAB VI .....                                                                                               | VI-1 |
| KESIMPULAN DAN SARAN .....                                                                                 | VI-1 |
| 6.1 Pendahuluan .....                                                                                      | VI-1 |
| 6.2 Kesimpulan .....                                                                                       | VI-1 |
| 6.3 Saran .....                                                                                            | VI-2 |
| DAFTAR PUSTAKA .....                                                                                       | xii  |

## DAFTAR TABEL

|                                                                                                                 |        |
|-----------------------------------------------------------------------------------------------------------------|--------|
| Tabel II-1. Skema 10 fold cross validation .....                                                                | II-9   |
| Tabel II-2. Confusion Matrix binary class .....                                                                 | II-10  |
| Tabel III-1. Rancangan Tabel Hasil Model Klasifikasi Confusion Matrix .....                                     | III-5  |
| Tabel III-2. Rancangan Tabel Hasil Klasifikasi Menggunakan Algoritma Naïve Bayes dengan Forward Selection ..... | III-5  |
| Tabel III-5. Rancangan Tabel Hasil Analisis Klasifikasi.....                                                    | III-7  |
| Tabel III-6. Tabel Work Breakdown Structure (WBS) Dari Penelitian Yang Akan Dilakukan .....                     | III-11 |
| Tabel IV-1. Definisi Aktor Use Case.....                                                                        | IV-7   |
| Tabel IV-2. Definisi Use Case.....                                                                              | IV-8   |
| Tabel IV-3 Skenario Use Case memasukan data .....                                                               | IV-9   |
| Tabel IV-4. Skenario Use Case melakukan klasifikasi dengan naive bayes....                                      | IV-10  |
| Tabel IV-5. Skenario Use Case melakukan klasifikasi data baru.....                                              | IV-12  |
| Tabel IV-6. Implementasi Kelas.....                                                                             | IV-22  |
| Tabel IV-7. Rencana Pengujian Use Case Memuat Data .....                                                        | IV-24  |
| Tabel IV-8. Rencana Pengujian Use Case Melakukan Klasifikasi dengan Naive Bayes.....                            | IV-25  |
| Tabel IV-9. Rencana Pengujian Use Case Melakukan Klasifikasi dengan Naive Bayes dan forward selection .....     | IV-25  |
| Tabel IV-10. Pengujian Use Case Memuat Data .....                                                               | IV-26  |
| Tabel IV-11. Pengujian Use Case Melakukan Klasifikasi dengan Naive Bayes.....                                   | IV-27  |
| Tabel IV-12 Hasil Pengujian Tabel Confusion Matrix. ....                                                        | IV-28  |



|                                                                                                        |      |
|--------------------------------------------------------------------------------------------------------|------|
| Tabel V-1 Hasil Pengujian Tabel Confusion Matrix .....                                                 | V-2  |
| Tabel V-2 Hasil Pengujian Tabel Confusion Matrix .....                                                 | V-3  |
| Tabel V-3 Hasil Pengujian Tabel Confusion Matrix .....                                                 | V-4  |
| Tabel V-4 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-1 .....                              | V-5  |
| Tabel V-5 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-2 .....                              | V-6  |
| Tabel V-6 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-3 .....                              | V-7  |
| Tabel V-7 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-4 .....                              | V-8  |
| Tabel V-8 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-5 .....                              | V-9  |
| Tabel V-9 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-6 .....                              | V-10 |
| Tabel V-10 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-7 .....                             | V-10 |
| Tabel V-11 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-8 .....                             | V-11 |
| Tabel V-12. Tabel Hasil Klasifikasi Menggunakan Algoritma Naïve Bayes<br>dengan Forward Selection..... | V-12 |
| Tabel V-13 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-1 .....                             | V-14 |
| Tabel V-14 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-2 .....                             | V-15 |
| Tabel V-15 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-3 .....                             | V-16 |
| Tabel V-16 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-4 ....                              | V-17 |
| Tabel IV-17 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-5 ...                              | V-17 |
| Tabel V-18 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-6 ....                              | V-18 |
| Tabel V-19 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-7 ....                              | V-19 |
| Tabel V-20 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-8 ....                              | V-20 |
| Tabel V-21 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-9 ....                              | V-21 |
| Tabel V-22 Hasil Klasifikasi Menggunakan Algoritma Naïve Bayes dengan<br>Forward Selection.....        | V-21 |

|                                                                            |      |
|----------------------------------------------------------------------------|------|
| Tabel V-23 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-1 ..... | V-23 |
| Tabel V-24 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-2 ..... | V-24 |
| Tabel V-25 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-3 ..... | V-25 |
| Tabel V-26 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-4 ..... | V-26 |
| Tabel V-27 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-5 ..... | V-27 |
| Tabel V-28 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-6 ..... | V-27 |
| Tabel V-29 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-7 ..... | V-28 |
| Tabel V-30 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-8 ..... | V-29 |
| Tabel V-31 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-9 ..... | V-30 |
| Tabel V-32 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-10 .... | V-31 |
| Tabel V-33 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-11 .... | V-32 |
| Tabel V-34 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-12 .... | V-33 |
| Tabel V-35 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-13 .... | V-34 |
| Tabel V-36 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-14 .... | V-35 |
| Tabel V-37 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-15 .... | V-36 |
| Tabel V-38 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-16 .... | V-37 |
| Tabel V-39 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-17 .... | V-38 |
| Tabel V-40 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-18 .... | V-39 |
| Tabel V-41 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-19 .... | V-40 |
| Tabel V-42 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-20 .... | V-41 |
| Tabel V-43 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-21 .... | V-42 |
| Tabel V-44 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-22 .... | V-43 |
| Tabel V-45 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-23 .... | V-44 |

|                                                                                                        |      |
|--------------------------------------------------------------------------------------------------------|------|
| Tabel V-46 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-24 ....                             | V-45 |
| Tabel V-47 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-25 ....                             | V-46 |
| Tabel V-48 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-26 ....                             | V-47 |
| Tabel V-49 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-27 ....                             | V-48 |
| Tabel V-50 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-28 ....                             | V-49 |
| Tabel V-51 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-29 ....                             | V-50 |
| Tabel V-52 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-30 ....                             | V-51 |
| Tabel V-53 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-31 ....                             | V-52 |
| Tabel V-54 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-32 ....                             | V-53 |
| Tabel V-55 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-33 ....                             | V-54 |
| Tabel V-56 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-34 ....                             | V-55 |
| Tabel V-57 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-35 ....                             | V-56 |
| Tabel V-58 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-36 ....                             | V-57 |
| Tabel V-59 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-37 ....                             | V-58 |
| Tabel V-60 Hasil Pengujian Tabel Confusion Matrix Model Atribut ke-38 ....                             | V-59 |
| Tabel V-61. Tabel Hasil Klasifikasi Menggunakan Algoritma Naïve Bayes<br>dengan Forward Selection..... | V-60 |
| Tabel V-62. Tabel Hasil Analisis Klasifikasi.....                                                      | V-67 |
| Tabel V-63. Tabel Hasil Analisis Klasifikasi.....                                                      | V-69 |

## DAFTAR GAMBAR

|                                                                                                   |       |
|---------------------------------------------------------------------------------------------------|-------|
| Gambar III-1. Tahapan Pengujian Penelitian. ....                                                  | III-6 |
| Gambar III-2. Arsitektur RUP .....                                                                | III-8 |
| Gambar IV-1. Diagram Use Case .....                                                               | IV-7  |
| Gambar IV-2. Rancangan Antar Muka Perangkat Lunak .....                                           | IV-13 |
| Gambar IV-3 Rancangan Fitur Muat Data .....                                                       | IV-13 |
| Gambar IV-4. Rancangan Pemodelan Klasifikasi .....                                                | IV-14 |
| Gambar IV-5. Scroll-box yang akan menampilkan nilai accuracy, presicion dan recall ..             | IV-14 |
| Gambar IV-7. Sequence diagram melakukan klasifikasi dengan naive bayes .                          | IV-16 |
| Gambar IV-8 Sequence diagram melakukan klasifikasi dengan naive bayes dan forward selection ..... | IV-17 |
| Gambar IV-9. Activity diagram melakukan muat dokumen .....                                        | IV-18 |
| Gambar IV-10. Activity diagram melakukan klasifikasi dengan naive bayes .                         | IV-19 |
| Gambar IV-12. Diagram Kelas .....                                                                 | IV-21 |

# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab pendahuluan menjelaskan tentang pokok-pokok pikiran yang melandasi rencana skripsi. Pokok-pokok pikiran dimaksud antara lain latar belakang masalah penelitian, perumusan masalah/permasalahan penelitian, tujuan penelitian dan manfaat penelitian, dan batasan masalah.

Bab pendahuluan dimulai dengan penjelasan mengenai tantangan dan tujuan proses menemukan pengetahuan baru pada *Data Mining*, penelitian yang berkaitan dengan masalah, serta bagaimana penyelesaian yang akan dilakukan.

### **1.2 Latar Belakang**

Proses pengolahan data dengan jumlah besar untuk menghasilkan pola pengetahuan berdasarkan tugasnya terdapat empat kelas tugas, yaitu klustering untuk pengelompokan dimana kelas belum ditentukan sebelumnya, regresi untuk berusaha menemukan fungsi yang memodelkan data dengan kesalahan paling sedikit, asosiasi untuk mencari hubungan antar variabel, dan klasifikasi untuk mengatur data ke dalam kelompok yang telah ditentukan atau dengan kata lain proses menemukan seperangkat model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan untuk menggunakan model dalam memprediksi kelas objek yang label kelasnya tidak diketahui (Li, 2010).

Salah satu metode yang dapat digunakan dalam pengklasifikasian yaitu *Naive Bayes*. *Naive bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Dalam penelitian (Rathi & Singh, 2018), metode *naive bayes* dapat memprediksi penyakit kanker payudara dengan tingkat akurasi sebesar 61,30%, *recall* 39,10%, dan spesifitas 88,30%. Namun algoritma *naive bayes* mengasumsikan fitur yang independen. Teorema *Bayes* yang mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas disebut *naive* (Patil, & Sherekar, 2013). Fitur independen yaitu fitur yang berdiri sendiri atau tidak terpengaruh oleh variabel lain yang di ukur dalam penelitian serta tidak boleh dikorelasikan satu sama lain. Fitur dependen merupakan fitur yang memiliki pengaruh terhadap variabel lain yang di ukur dalam penelitian (Gogtay, Deshpande, & Thatte, 2017). Asumsi independen pada fitur di kehidupan nyata hampir tidak pernah terjadi (Zhang, 2004). Sehingga untuk mengatasi masalah asumsi fitur independen dalam pengoptimasian *naive bayes* dapat menggunakan metode fitur seleksi.

Fitur seleksi merupakan metode untuk memilih fitur atau variabel yang memiliki beberapa manfaat seperti memfasilitasi visualisasi data, pemahaman data, dan meningkatkan kinerja prediksi dalam hal mempercepat serta memaksimalkan kerja suatu algoritma (Guyon, 2003). Salah satu metode fitur seleksi yaitu metode *Forward Selection*. *Forward selection* akan menghapus variabel yang tidak relevan dari model prediksi dengan cara regresi linear (Luo & Ghosal, 2016), dimana regresi linear berfokus pada asumsi ketergantungan atau sebab akibat hubungan

antara satu atau lebih variabel independen dan variabel dependen (Gogtay et al., 2017). Untuk mengetahui apakah memiliki hubungan atau tidak antara variabel yaitu dengan menggunakan metode *Pearson Correlation*.

Penelitian yang akan dilakukan bertujuan untuk mengetahui pengaruh metode seleksi fitur dalam hal ini menggunakan metode *forward selection* pada klasifikasi algoritma *naïve bayes*. Dimana korelasi antar atribut diperhitungkan terlebih dahulu menggunakan metode *pearson correlation*. Lalu dengan metode *forward selection* untuk menentukan sebuah model fitur yang paling baik. Dalam penelitian ini, akan dilakukan perbandingan hasil klasifikasi antara algoritma *naïve bayes* tanpa metode seleksi fitur *forward selection*, dengan algoritma *naïve bayes* yang dikombinasi dengan metode seleksi fitur *forward selection* dalam pengklasifikasian kemudian di evaluasi menggunakan *confusion matrix* untuk mendapatkan nilai *accuracy* yaitu kedekatan antara nilai prediksi dan aktual, nilai *recall* yaitu keberhasilan sistem dalam menjawab informasi, dan nilai *precision* yaitu ketepatan sistem dalam menjawab informasi.

### **1.3 Rumusan Masalah**

Rumusan masalah dari penelitian ini adalah bagaimana pengaruh *Forward Selection* pada klasifikasi menggunakan *naïve bayes*? Untuk menjawab rumusan masalah tersebut, diuraikan beberapa pertanyaan penelitian sebagai berikut :

1. Bagaimana mekanisme metode *forward selection* dalam melakukan seleksi fitur?

2. Bagaimana mekanisme metode *naïve bayes* dalam melakukan pengklasifikasian?
3. Bagaimana nilai *accuracy*, *recall*, dan *precision* pengklasifikasian menggunakan *naïve bayes* dengan adanya seleksi fitur menggunakan *forward selection*?
4. Bagaimana nilai *accuracy*, *recall*, dan *precision* pengklasifikasian menggunakan *naïve bayes* tanpa seleksi fitur *forward selection*?

#### **1.4 Tujuan Penelitian**

Tujuan dari penelitian yang dilakukan adalah sebagai berikut :

1. Memahami mekanisme metode *forward selection* dalam melakukan seleksi fitur
2. Memahami mekanisme metode *naïve bayes* dalam melakukan pengklasifikasian
3. Mengetahui nilai *accuracy*, *recall*, dan *precision* pengklasifikasian menggunakan *naïve bayes* dengan adanya seleksi fitur menggunakan *forward selection*
4. Mengetahui nilai *accuracy*, *recall*, dan *precision* pengklasifikasian menggunakan *naïve bayes* tanpa seleksi fitur *forward selection*

#### **1.5 Manfaat Penelitian**

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Memahami *forward selection* sebagai metode seleksi fitur



2. Memahami algoritma *naïve bayes* dikombinasi dengan metode *forward selection* dalam pengklasifikasian

## **1.6 Batasan Masalah**

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Dataset yang digunakan berupa dataset yang diperoleh dari publik repositori UCI dan Kaggle yaitu *Pima Indians Diabetes, Breast Cancer Coimbra, & SPECTF Heart*.
2. Metode-metode yang digunakan yaitu metode klasifikasi *Naive Bayes* dan metode fitur seleksi *Forward Selection*.

## **1.7 Sistematika Penulisan**

### **BAB I. PENDAHULUAN**

Pada bab pendahuluan, menguraikan gagasan yang melandasi pembuatan tugas akhir yaitu mengenai latar belakang masalah, perumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan masalah.

### **BAB II. KAJIAN LITERATUR**

Pada bab ini membahas dasar-dasar teori yang digunakan pada penelitian. Teori yang akan diuraikan pada bab ini seperti

definisi dan algoritma *naïve bayes*, *forward selection*, *confusion matrix* dan penelitian yang terkait pada penelitian ini.

### **BAB III. METODELOGI PENELITIAN**

Bab ini menguraikan tahapan tahapan yang akan dilakukan dalam penelitian ini. Setiap tahapan yang akan diuraikan secara rinci dan mengacu pada suatu kerangka kerja.

### **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Pada bab ini membahas mengenai analisis dan perancangan perangkat lunak yang akan digunakan sebagai alat penelitian. Dimulai dari pengumpulan dan analisa kebutuhan, rancangan dan konstruksi perangkat lunak serta pengujian untuk memastikan semua kebutuhan pengembangan perangkat lunak sesuai dengan dengan kebutuhan. Penyusunan pada bab ini memiliki kerangka penulisan dengan fase-fase dan elemen-elemen pengembangan perangkat lunak bersifat berorientasi objek.

### **BAB V. HASIL DAN ANALISA PENELITIAN**

Pada bab ini diuraikan hasil pengujian berdasarkan langkah-langkah yang telah direncanakan. Tabel hasil pengujian serta analisisnya disajikan sebagai basis dari kesimpulan yang akan diambil dalam penelitian ini.

## **BAB VI. KESIMPULAN DAN SARAN**

Pada bab ini berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga saran-saran yang diharapkan berguna dalam melakukan pengklasifikasian dengan metode klasifikasi *naïve bayes*.

## **1.8 Kesimpulan**

Dari bab pendahuluan ini telah dijelaskan bahwa penelitian ini meliputi latar belakang dan rumusan masalah, tujuan dan manfaat penelitian, batasan masalah dan sistematika penulisan. Serta penelitian yang akan di lakukan yaitu untuk mengetahui pengaruh metode seleksi *forward selection* terhadap pengklasifikasian *naïve bayes* dan model yang dihasilkan. Kemudian hasil tingkat *accuracy*, *recall*, dan *precision* dihitung menggunakan *confusion matrix*.

## DAFTAR PUSTAKA

- Ammar, S. M. (2009). Keoptimalan Naive Bayes Dalam Klasifikasi.
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering*, 40(1), 16–28.  
<https://doi.org/10.1016/j.compeleceng.2013.11.024>
- FITRIANI, R. Y. (2015). *PENGARUH KOMUNIKASI INTERPERSONAL DAN GAYA KEPEMIMPINAN TERHADAP EFEKTIVITAS KERJA PEGAWAI PADA KANTOR KEPALA DESA CATURHARJO KECAMATAN SLEMAN KABUPATEN SLEMAN*.
- Gogtay, N. J., Deshpande, S. P., & Thatte, U. M. (2017). *Principles of Regression Analysis*. 65(April), 48–52.
- Guyon, I. (2003). *An Introduction to Variable and Feature Selection I Introduction*. 3, 1157–1182.
- Hong, H. G., Zheng, Q., & Li, Y. (2019). *Forward regression for Cox models with high-dimensional covariates*. 173, 268–290. <https://doi.org/10.1016/j.jmva.2019.02.011>
- Kapourani, A. C. (2018). *Naive Bayes & Introduction to Gaussians*. 1–10.
- Kumar, S. V. K., & Kiruthika, P. (2015). An Overview of Classification Algorithm in Data mining. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(12), 255–257. <https://doi.org/10.17148/IJARCCE.2015.41259>
- Luo, S., & Ghosal, S. (2016). Forward selection and estimation in high dimensional single index models. *Statistical Methodology*, 33, 172–179.  
<https://doi.org/10.1016/j.stamet.2016.09.002>
- Li, Y. (2010). *Data Mining: Concepts, Background and Methods of Integrating Uncertainty*

in Data Mining. *CCSC:SC Student E-Journal*, 3, 2–7.

<https://doi.org/10.1017/S1092852900015340>

Luo, S., & Ghosal, S. (2016). Forward selection and estimation in high dimensional single index models. *Statistical Methodology*, 33, 172–179.

<https://doi.org/10.1016/j.stamet.2016.09.002>

Nuruddin, S. A. (2017). *ALGORITMA K-NEAREST NEIGHBOR BERBASIS FORWARD SELECTION UNTUK MENDIAGNOSIS PENYAKIT JANTUNG*. 3(1).

Ramsay, B., & Knaap, E. Van Der. (2011). *Confusion Matrix-based Feature Selection Sofia Visa*.

Rathi, M., & Singh, A. K. (2018). *Breast Cancer Prediction using Naïve Bayes Classifier*. *Breast Cancer Prediction using Naïve Bayes Classifier*. (July 2012).

Sedgwick, P., & George, S. (2016). *Pearson's correlation coefficient*. (January), 3–5.

<https://doi.org/10.1136/bmj.e4483>

Zhang, H. (2004). *The Optimality of Naive Bayes*