TESIS

KLASIFIKASI KELULUSAN MAHASISWA MENGGUNAKAN METODE *NAÏVE BAYES* DENGAN KOMPARASI PRAPEMROSESAN *RANDOM OVERSAMPLING* SERTA SELEKSI FITUR *INFORMATION GAIN* DAN *FORWARD SELECTION*

(Studi Kasus: Fakultas Sains dan Teknologi UIN SUSKA Riau)

Oleh:

Dony Fahrudy

NIM: 20206052006

PROGRAM STUDI INFORMATIKA

PROGRAM MAGISTER

FAKULTAS SAINS DAN TEKNOLOGI

UIN SUNAN KALIJAGA

YOGYAKARTA

2022

LEMBAR PENGESAHAN

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1764/Un.02/DST/PP.00.9/08/2022

Tugas Akhir dengan judul : KLASIFIKASI KELULUSAN MAHASISWA MENGGUNAKAN METODE NAIVE BAYES DENGAN KOMPARASI PRAPEMROSESAN RANDOM OVERSAMPLING SERTA SELEKSI FITUR INFORMATION GAIN DAN FORWARD SELECTION (STUDI KASUS; FAKULTAS SAINT DAN TEKNOLOGI UIN SUSKA RIAU)

yang dipersiapkan dan disusun oleh:

Nama : DONY FAHRUDY, S.T.
Nomor Induk Mahasiswa : 20206052006
Telah diujikan pada : Senin, 25 Juli 2022
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR

Ketua Sidang

Dr. Ir. Shofwatul 'Uyun, S.T., M.Kom.
SIGNED

Valid ID: 62fb6981c95a7

Penguji I

Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.
SIGNED

Valid ID: 62fdd349c8fc2

Penguji II

Dr. Ir. Bambang Sugiantoro, S.Si., M.T.
SIGNED

Valid ID: 62e0a8ee7e2af

Yogyakarta, 25 Juli 2022
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 62fdef282d1ef

1/1 18/08/2022

PERNYATAAN KEASLIAN / BEBAS PLAGIASI

SURAT PERNYATAAN KEASLIAN / BEBAS PLAGIASI

Yang bertanda tangan dibawah ini:

Nama Mahasiswa	: Dony Fahrudy
NIM	: 20206052006
Program Studi	: Informatika (S2)
Fakultas	: Sains dan Teknologi

Menyatakan dengan sesungguhnya bahwa laporan tesis saya yang berjudul: **"Klasifikasi Kelulusan Mahasiswa Menggunakan Metode Naïve Bayes dengan Komparasi Prapemrosesan Random Oversampling serta Seleksi Fitur Information Gain dan Forward Selection (Studi Kasus: Fakultas Sains dan Teknologi UIN SUSKA Riau)"** adalah hasil karya pribadi yang tidak mengandung plagiarisme dan tidak berisi materi yang dipublikasikan atau ditulis orang lain, kecuali bagian – bagian tertentu yang penulis ambil sebagai acuan dan tata cara yang diberikan secara ilmiah.

Jika terbukti pernyataan ini tidak benar, maka penulis siap mempertanggung jawabkan sesuai hukum yang berlaku.

Yogyakarta, 14 Juli 2022

Yang menyatakan,

Dony Fahrudy

NIM. 20206052006

HALAMAN PERSETUJUAN

Universitas Islam Negeri Sunan Kalijaga

SURAT PERSETUJUAN TUGAS AKHIR

Hal : Persetujuan Tugas Akhir
Lamp : -

Kepada Yth.,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta

Assalamualaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka saya selaku pembimbing berpendapat bahwa naskah tesis Saudara:

Nama	:	Dony Fahrudy
NIM	:	20206052006
Judul Tesis	:	Klasifikasi Kelulusan Mahasiswa Menggunakan Metode *Naïve Bayes* dengan Komparasi Prapemrosesan Random Oversampling serta Seleksi Fitur Information Gain dan Forward Selection (Studi Kasus: Fakultas Sains dan Teknologi UIN SUSKA Riau)

Saya berpendapat bahwa tesis tersebut sudah dapat diajukan kepada Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta untuk diujikan dalam rangka memperoleh gelar Magister Informatika.

Waalaikumsalam wr. wb.

Yogyakarta, 14 Juli 2022
Pembimbing,

Dr. Ir. Shofwatul 'Uyun, S.T., M.Kom.
NIP. 19820511 200604 2 002

ABSTRAK

Ketidakseimbangan jumlah data pada setiap kelas serta memiliki dimensi atribut tinggi pada *datasets*, seringkali menjadi masalah dalam proses klasifikasi yang dapat mempengaruhi kinerja algoritma dalam proses komputasi, karena terdapat jumlah data pada setiap kelas yang tidak seimbang dan atribut yang tidak relevan harus diproses, sehingga perlunya dilakukan teknik mengatasi ketidakseimbangan kelas untuk menyeimbangkan jumlah data pada setiap kelas dan seleksi fitur untuk mengurangi kompleksitas data dan fitur-fitur yang tidak relevan. Oleh karena itu, penelitian ini menggunakan teknik *Random Oversampling* (ROs) untuk mengatasi ketidakseimbangan kelas, serta dua metode seleksi fitur dengan algoritma *Information Gain* dan *Forward Selection* yang dikomparasikan untuk menentukan metode seleksi fitur yang lebih unggul, lebih efektif, dan lebih cocok digunakan. Hasil dari seleksi fitur digunakan untuk melakukan klasifikasi kelulusan mahasiswa dengan membangun model klasifikasi menggunakan algoritma *Naïve Bayes*. Hasil penelitian menunjukkan peningkatan rata-rata akurasi dari metode *Naïve Bayes* dengan tanpa prapemrosesan ROs dan seleksi fitur, penggunaan ROs, penggunaan *Information Gain* dengan 3 fitur terpilih serta *Forward Selection* dengan 2 fitur terpilih secara berurutan adalah 81.83%; 83.84%; 86.03% dan 86.42%, sehingga terjadi peningkatan akurasi sebesar 4.2% dari tanpa prapemrosesan ke *Information Gain* dan 4.59% dari tanpa prapemrosesan ke *Forward Selection*. Oleh karena itu, metode seleksi fitur terbaik adalah *Forward Selection* dengan 2 fitur terpilih (Ip Semester 8 dan IPK), penggunaan ROs dan kedua seleksi fitur terbukti meningkatkan kinerja metode *Naïve Bayes*.

Kata Kunci: *Forward Selection, Information Gain,* Kelulusan Mahasiswa, *Naïve Bayes, ROs*

ABSTRACT

The imbalance in the amount of data in each class and having high attribute dimensions in datasets, is often a problem in the classification process that can affect the algorithm's performance in the computational process, because there is an unbalanced amount of data in each class and irrelevant attributes must be processed, so it is necessary to do this. class imbalance techniques to balance the amount of data in each class and feature selection to reduce data complexity and irrelevant features. Therefore, this study uses the Random Oversampling (ROs) technique to overcome class imbalances, as well as two feature selection methods with Information Gain and Forward Selection algorithms which are compared to determine which feature selection method is superior, more effective, and more suitable. used. The results of feature selection are used to classify student graduation by building a classification model using the Naïve Bayes algorithm. The results showed an increase in the average accuracy of the Naïve Bayes method with no ROs preprocessing and feature selection, the use of ROs, the use of Information Gain with 3 selected features and Forward Selection with 2 selected features sequentially was 81.83%; 83.84%; 86.03% and 86.42%, so that there is an increase in accuracy of 4.2% from no pre-processing to Information Gain and 4.59% from no pre-processing to Forward Selection. Therefore, the best feature selection method is Forward Selection with 2 selected features (Ip Semester 8 and GPA), the use of ROs and both feature selections are proven to improve the performance of the Naïve Bayes method.

***Keywords** : Forward Selection, Information Gain, Student Graduation, Naïve Bayes, ROs*

MOTTO

“Sesungguhnya Allah tidak akan mengubah keadaan suatu kaum, sebelum mereka mengubah keadaan diri mereka sendiri.”
(QS Ar Rad 11)

“Rahasia kesuksesan adalah mengetahui yang orang lain belum ketahui.”
(Aristotle Onassis)

“Pengetahuan yang baik adalah yang memberikan manfaat, bukan hanya diingat.”
(Imam Syafi’i)

“Hari esok harus lebih baik dari hari ini.”
(Dony Fahrudy)

KATA PENGANTAR

Puji dan Syukur Kehadirat Allah SWT atas segala rahmat dan karunia-Nya, akhirnya penulis dapat menyelesaikan penyusunan tesis yang berjudul "Klasifikasi Kelulusan Mahasiswa Menggunakan Metode *Naïve Bayes* dengan Komparasi Prapemrosesan *Random Oversampling* serta Seleksi Fitur *Information Gain* dan *Forward Selection* (Studi Kasus : Fakultas Sains dan Teknologi UIN SUSKA Riau)".

Tesis ini ditulis dalam rangka memenuhi persyaratan untuk memperoleh gelar Magister di UIN Sunan Kalijaga Yogyakarta. Penulis menyadari bahwa tesis dapat diselesaikan berkat dukungan dan bantuan dari berbagai pihak, oleh karena itu penulis berterima kasih kepada semua pihak yang secara langsung maupun tidak langsung memberikan kontribusi dalam menyelesaikan Tesis ini. Selanjutnya ucapan terima kasih penulis sampaikan kepada :

1. Bapak Dr. Phil. Sahiron, M.A. sebagai Plt. Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Dr. Khurul Wardati, M,Si, selaku dekan fakultas SAINTEK UIN Sunan Kalijaga.
3. Bapak Dr. Bambang Sugiantoro, S.Si., M.T. sebagai Kepala Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sekaligus penguji kedua, yang telah memberikan izin dan kemudahan, sehingga penulis dapat menyelesaikan studi di Program Studi Program Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
4. Ibu Dr. Ir. Shofwatul 'Uyun, S.T., M.Kom. sebagai pembimbing akademik sekaligus pembimbing tugas akhir, dengan kesibukannya masih tetap meluangkan waktunya untuk memberikan bimbingan, petunjuk, dan mendorong semangat penulis untuk menyelesaikan penulisan Tesis ini.
5. Ibu Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D, M.Kom. sebagai penguji pertama, dengan kesibukannya masih tetap meluangkan waktunya untuk memberikan petunjuk, saran, masukan dan mendorong semangat penulis untuk

menyelesaikan penulisan Tesis ini.

6. Seluruh dosen dan staf administrasi pada Program Studi Informatika Program Magister Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta, yang secara langsung atau tidak langsung telah memberi bantuan kepada penulis dalam menyelesaikan penulisan tesis.
7. Segenap staf UPT. Pusat Teknologi Informasi dan Pangkalan Data UIN Sunan Kalijaga Yogyakarta yang secara langsung atau tidak langsung telah memberi bantuan kepada penulis dalam menyelesaikan penulisan tesis.
8. Orang Tua tercinta yang mendidik dengan penuh rasa kasih sayang dan senantiasa memberi semangat dan dorongan kepada penulis.
9. Keluarga tercinta yang telah memberikan do'a dan dukungan kepada penulis.
10. Seluruh rekan-rekan mahasiswa di Magister Informatika yang telah saling mendukung untuk melalui perjuangan bersama-sama.

Akhirnya penulis berharap semoga Tesis ini dapat bermanfaat dan permintaan maaf yang tulus jika seandainya dalam penulisan ini terdapat kekurangan dan kekeliruan, penulis juga menerima kritik dan saran yang bersifat membangun demi menyempurnakan penulisan Tesis ini.

Yogyakarta, 25 Juli 2022

Penulis

Dony Fahrudy

NIM. 20206052006

DAFTAR ISI

DAFTAR TABEL

Tabel **Halaman**

DAFTAR GAMBAR

Gambar **Halaman**

DAFTAR LAMPIRAN

Lampiran **Halaman**

BAB I
PENDAHULUAN

A. Latar Belakang

Kelulusan mahasiswa merupakan faktor keberhasilan yang menjadi salah satu penilaian akreditasi pada perguruan tinggi. Apabila mahasiswa lulus tepat waktu dalam menyelesaikan studinya, maka dapat mendukung penilaian akreditasi tersebut (ACCJC/WASC, 2012). Salah satu kriteria keberhasilan mahasiswa dalam memperoleh gelar sarjana pada perguruan tinggi adalah lulus tepat waktu, sehingga mahasiswa dapat lulus tepat waktu apabila telah menyelesaikan studinya selama kurang dari atau sama dengan empat tahun. Namun dalam pelaksanaannya mahasiswa tidak selalu dapat menyelesaikan pendidikan sarjana selama kurang dari atau sama dengan empat tahun (Nuffic, 2017).

Berdasarkan *datasets* yang diperoleh pada perguruan tinggi, mahasiswa tidak selalu dapat menyelesaikan pendidikan sarjana secara tepat waktu, sehingga diperlukan peran perguruan tinggi untuk mengantisipasi hal tersebut. Otoritas pendidikan, administrasi akademik, dan orang tua prihatin dengan penurunan tingkat kelulusan mahasiswa di lembaga pendidikan tinggi. Pihak perguruan tinggi mencoba untuk meningkatkan kelulusan mahasiswa. Kemampuan klasifikasi untuk mengantisipasi kelulusan mahasiswa secara akurat sangat penting karena memungkinkan lembaga pendidikan untuk menetapkan program strategis untuk membantu dan meningkatkan kinerja mahasiswa menuju kelulusan tepat waktu (Bassi et al., 2019). Oleh karena itu, perlunya peningkatan kualitas akademik oleh institusi dan mengoptimalkan sumber daya agar dapat membantu mahasiswa menyelesaikan studi mereka secara tepat waktu (Lei & Li, 2015). Oleh karena itu, diperlukan sebuah model untuk mengklasifikasi kelulusan mahasiswa yang merupakan langkah terpenting untuk mengetahui kualitas mahasiswa (Bassi et al., 2019). Ketersediaan data latih dan uji untuk setiap kelas merupakan salah satu kriteria yang menentukan keberhasilan suatu model. Ada beberapa isu yang seringkali ditemukan dalam komputasi, yaitu: jumlah data antar kelas yang tidak

seimbang dan data dengan dimensi atribut tinggi.

Masalah ketidakseimbangan kelas dapat terjadi ketika *instance* dari satu kelas melebihi jumlah *instance* (kelas mayoritas) dari kelas lain (kelas minoritas), sehingga menyebabkan kesalahan klasifikasi pada kelas minoritas yang berdampak bias pada hasil klasifikasi kelas mayoritas. Secara umum, ada tiga teknik *resampling* yaitu *random undersampling*, *random oversampling* dan *hybrid methods* (Q. Wang et al., 2017). Untuk mengatasi masalah ketidakseimbangan pada kelas minoritas dapat menggunakan teknik *random oversampling* dengan mereplikasi *instance* pada kelas minoritas secara acak (Naganjaneyulu & Kuppa, 2013).

Pada umumnya algoritma klasifikasi menggunakan semua fitur pada data untuk membangun sebuah model, namun tidak semua fitur tersebut relevan untuk proses klasifikasi. Jika hal tersebut terdapat pada data yang memiliki ukuran dan dimensi yang sangat besar, maka dapat membuat kinerja algoritma menjadi tidak efektif akibat banyak fitur yang tidak relevan harus diproses. Salah satu solusi yang dapat mengatasi masalah tersebut adalah dengan menggunakan seleksi fitur. Seleksi fitur merupakan salah satu tahap dalam *preprocessing* pada klasifikasi yang dilakukan dengan cara memilih atribut yang relevan terhadap data yang mempengaruhi hasil klasifikasi. Seleksi fitur juga digunakan untuk meningkatkan efektifitas dan efisiensi kinerja dari algoritma klasifikasi, serta dapat mengurangi dimensi data dan fitur yang tidak relevan (Vanaja & Ramesh Kumar, 2014). Tujuan dari seleksi fitur adalah untuk mengurangi tingkat kompleksitas dari sebuah algoritma klasifikasi, meningkatkan akurasi dari algoritma klasifikasi, dan dapat mengetahui atribut yang mempengaruhi kinerja algoritma (C.Deisy et al., 2007). Secara umum, dasar ide algoritma *feature selection* yaitu mencari semua kemungkinan kombinasi dari atribut dalam data untuk menemukan subset dari feature yang terbaik untuk klasifikasi (Ting et al., 2011).

Terdapat tiga pendekatan dalam seleksi fitur, yaitu *filter*, *wrapper* dan *embedded* (Vanaja & Ramesh Kumar, 2014). Metode *filter* merupakan metode seleksi fitur yang memilih variabel berdasarkan peringkat dan menghilangkan variabel yang berada di bawah ambang batas (Chandrashekar & Sahin, 2014).

Metode ini menggunakan kriteria teknik *ranking* yang sederhana dan menghasilkan fitur yang relevan. Seleksi fitur pada pendekatan *filter* menggunakan teknik evaluasi yang terpisah dengan algoritma pembelajaran (Visalakshi & Radha, 2015). Kelebihan dari metode *filter* adalah lebih cepat, terukur untuk kumpulan data yang besar, dan independen, selain itu teknik *filter* mudah diterapkan (Mwadulo, 2016). Kekurangan dari metode *filter* adalah dapat mengabaikan ketergantungan fitur dan kurangnya interaksi dengan *classifier* karena dilakukan secara terpisah (Ang et al., 2016).

Subset fitur yang dihasilkan metode *wrapper* menggunakan sekumpulan kombinasi fitur dari teknik pencarian kemudian melatih model prediktif pada himpunan bagian dari fitur dan mengevaluasi *subset* tersebut menggunakan algoritma pembelajaran terawasi untuk mengetahui akurasi kinerja. Setiap kombinasi fitur dibandingkan satu dengan yang lainnya dan digunakan model algoritma untuk mengevaluasi kumpulan kombinasi yang ada (Gnana et al., 2016). Metode *wrapper* mengatasi masalah dalam metode *filter*. Metode *wrapper* dapat berinteraksi dengan model pembelajaran *classifier* dan memperhatikan keterkaitan antar atribut (Mwadulo, 2016). Metode *wrapper* dioptimalkan bersamaan dengan menggunakan algoritma pembelajaran klasifikasi. Oleh karena itu, secara umum, metode ini menghasilkan akurasi kinerja yang lebih baik daripada metode *filter*. Namun, metode ini secara komputasi mahal dan lebih kompleks daripada metode *filter* dan cenderung memerlukan waktu pemrosesan lebih lama (Jain & Singh, 2018).

Metode *embedded* mencari subset fitur yang optimal untuk algoritma klasifikasi tertentu ketika membangun sebuah *classifier* (Khaire & Dhanalakshmi, 2019). Metode ini membuat keputusan tergantung pada pengklasifikasi, sehingga pilihan fitur dipengaruhi oleh hipotesis pengklasifikasi dan tidak cocok dengan beberapa pengklasifikasi lainnya (Jain & Singh, 2018). Oleh karena itu, penelitian ini mengkomparasikan metode seleksi fitur dengan pendekatan *filter* dan *wrapper* dengan algoritma *information gain* dan *forward selection*.

Information gain adalah pendekatan pembelajaran mesin yang umum digunakan sebagai kriteria pemilihan atribut. *Information gain* secara khusus

memberikan *ranking* pada setiap *feature* (atribut) yang ada dan mengesampingkan *feature* (atribut) yang tidak memenuhi standar tertentu yang disusun berdasarkan nilai tertinggi ke terendah (Khemphila & Boonjing, 2011). Fitur dengan nilai *information gain* yang tinggi lebih bagus daripada fitur lainnya, yang mencerminkan bahwa semakin banyak informasi atribut yang berhubungan dengan kelas. *Information gain* tidak menghapus fitur yang berlebihan (Wah et al., 2018).

Forward selection merupakan algoritma pencarian paling sederhana. *Forward selection* adalah salah satu teknik untuk mereduksi dimensi *dataset* dengan menghilangkan atribut-atribut yang tidak relevan atau redundan (Panthong & Srivihok, 2015). Metode *forward selection* merupakan pemodelan yang diawali dengan nol *variable* (*empty model*) atau tidak ada *variable* dalam model, selanjutnya *variable* dimasukkan satu persatu. Untuk setiap *variable* yang ditambahkan, kinerja akan dievaluasi. Hanya atribut yang memiliki kinerja tertinggi ditambahkan ke seleksi untuk fungsi objek hingga kriteria tertentu dipenuhi (Kalyani & Karnan, 2010). Hasil dari seleksi fitur digunakan untuk membangun model klasifikasi menggunakan metode *data mining* dengan algoritma *naïve bayes* yang bertujuan untuk mengetahui teknik pemilihan fitur terbaik, lebih efisien dan lebih tepat untuk digunakan.

Menurut (Vanaja & Ramesh Kumar, 2014) *Data mining* merupakan kegiatan yang dilakukan berupa mengumpulkan data berukuran besar kemudian data tersebut diekstraksi menjadi sebuah pengetahuan (*knowledge*) yang dapat digunakan. Dalam *data mining*, terdapat beberapa metode yang dapat digunakan salah satunya adalah klasifikasi. Klasifikasi merupakan metode *data mining* yang digunakan untuk mendapatkan kelas yang belum diketahui yang dapat digunakan untuk memprediksi suatu kelas atau label tertentu.

Klasifikasi adalah suatu tipe data yang dianalisis yang dapat membantu orang menentukan kelas label dari sampel data yang ingin di klasifikasi dan mencari hubungan antara fitur masukan (*input*) dan fitur target (kelas) (Wah et al., 2018). Klasifikasi banyak digunakan untuk memprediksi suatu kelas atau label tertentu, yaitu dengan mengklasifikasi data dengan cara membangun model *berdasarkan data training* (data latih) dan memprediksi kelas atau label baru yang tidak

diketahui dari dataset yang mempunyai kelas menggunakan model dari klasifikasi untuk memprediksi data yang baru (Seh, 2019). Penelitian yang telah dilakukan oleh (Ashari et al., 2013) tentang perbandingan kinerja *Decision Tree*, *Naïve Bayes*, dan *K-Nearest Neighbor* adalah tiga teknik klasifikasi untuk mencari alternatif desain dalam alat replikasi energi. Hasil penelitian menunjukkan bahwa algoritma klasifikasi *Naïve Bayes* memiliki kinerja terbaik berdasarkan *Percision, Recall, Fmeasure, Accuracy,* dan AUC. *Naïve Bayes* mengungguli *Decision Tree* dan *K-Nearest Neighbor* pada semua parameter kecuali presisi. Oleh karena itu, algoritma klasifikasi yang digunakan untuk klasifikasi adalah algoritma *Naïve Bayes*.

Algoritma *Naïve Bayes* merupakan salah satu metode machine learning yang menggunakan perhitungan probabilitas dan statistik untuk memprediksi probabilitas atau peluang keanggotaan suatu class (Artaye, 2015). *Thomas Bayes* merupakan seorang ilmuwan Inggris yang menciptakan metode *Naïve Bayes*. Teorema Bayes menyatakan bahwa *Naïve Bayes* dapat memprediksi probabilitas masa depan berdasarkan data historis (Ali et al., 2021). Algoritma *Naïve Bayes* memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar (Artaye, 2015).

Fakultas Sains dan Teknologi merupakan salah satu fakultas yang ada di Universitas Islam Negeri Sultan Syarif Kasim Riau di Pekanbaru. *Datasets* diperoleh terdiri dari 4 jurusan yaitu Teknik Industri, Teknik Elektro, Sistem Informasi dan Matematika menunjukkan bahwa banyak mahasiswa yang menyelesaikan studinya dengan status tidak tepat waktu dapat berpengaruh pada penilaian akreditasi, sehingga perlu diperhatikan oleh perguruan tinggi dalam menentukan kualitas perguruan tinggi yang menjadi penilaian akreditasi. Pihak fakultas berusaha untuk membantu tenaga pengajar dalam menganalisa cara belajar, mendeteksi mahasiswa yang membutuhkan dukungan lebih. Hal tersebut perlu dilakukan untuk mencegah kegagalan akademik mahasiswa. Berdasarkan permasalahan tersebut, maka dalam hal ini peneliti akan melakukan penelitian tugas akhir tesis yang berjudul "Klasifikasi Kelulusan Mahasiswa Menggunakan Metode *Naïve Bayes* dengan Komparasi Prapemrosesan *Random Oversampling* serta Seleksi Fitur *Information Gain* dan *Forward Selection* (Studi Kasus : Fakultas Sains dan

Teknologi UIN SUSKA Riau)". Oleh karena itu, kontribusi penting dari penelitian ini adalah mengklasifikasi kelulusan mahasiswa untuk mengetahui kinerja akademik mahasiswa, menganalisis berbagai fitur untuk mengetahui fitur-fitur relevan untuk analisis data klasifikasi, selanjutnya dapat mengetahui perbandingan seleksi fitur mana yang terbaik dalam melakukan klasifikasi kelulusan mahasiswa serta mengetahui perbandingan kinerja algoritma *naïve bayes* sebelum dan setelah diterapkan prapemrosesan *random oversampling* dan seleksi fitur, selanjutnya diharapkan dapat membantu perguruan tinggi dan fakultas untuk mengetahui variabel yang berpengaruh terhadap tingkat kelulusan mahasiswa.

B. Rumusan Masalah

Berdasarkan permasalahan yang dijelaskan pada latar belakang, maka dapat dirumuskan masalah yang akan dijelaskan pada laporan Tugas Akhir Tesis ini adalah sebagai berikut :

1. Bagaimana menerapkan teknik *random oversampling* serta seleksi fitur *information gain* dan *forward selection* pada algoritma *naïve bayes* untuk klasifikasi kelulusan mahasiswa di Fakultas Sains dan Teknologi UIN SUSKA Riau ?
2. Bagaimana mengevaluasi perbandingan kinerja akurasi dalam penerapan teknik *random oversampling* pada algoritma *naïve bayes*, seleksi fitur *information gain* pada algoritma *naïve bayes*, seleksi fitur *forward selection* pada algoritma *naïve bayes* dan tanpa prapemrosesan *random oversampling* dan seleksi fitur untuk klasifikasi kelulusan mahasiswa di Fakultas Sains dan Teknologi UIN SUSKA Riau ?

C. Batasan Masalah

Ada beberapa batasan masalah dalam penelitian ini yaitu sebagai berikut :

1. Data yang dibutuhkan sebagai *datasets* utama adalah data akademik mahasiswa Fakultas Sains dan Teknologi UIN SUSKA Riau dari tahun 2016-2019 yang sudah dinyatakan lulus sebanyak 1420 data.

2. Kelas atau label yang digunakan sebagai hasil klasifikasi yaitu tepat waktu dan tidak tepat waktu.

D. Tujuan Penelitian

Tujuan dalam penelitian ini adalah sebagai berikut :

1. Menerapkan teknik *random oversampling* serta seleksi fitur *information gain* dan *forward selection* pada algoritma *naïve bayes* untuk klasifikasi kelulusan mahasiswa di Fakultas Sains dan Teknologi UIN SUSKA Riau.
2. Mengevaluasi perbandingan kinerja akurasi dalam penerapan teknik *random oversampling* pada algoritma *naïve bayes*, seleksi fitur *information gain* pada algoritma *naïve bayes*, seleksi fitur *forward selection* pada algoritma *naïve bayes* dan tanpa prapemrosesan *random oversampling* dan seleksi fitur untuk klasifikasi kelulusan mahasiswa di Fakultas Sains dan Teknologi UIN SUSKA Riau.

E. Manfaat Penelitian

Adapun manfaat penelitian adalah :

1. Dengan menganalisis berbagai fitur dapat mengetahui fitur-fitur relevan, terutama untuk analisis data klasifikasi.
2. Dengan adanya penelitian komparasi teknik *random oversampling* serta seleksi fitur *information gain* dan *forward selection* pada algoritma *naïve bayes* untuk klasifikasi kelulusan mahasiswa, dapat mengetahui perbandingan seleksi fitur mana yang terbaik dalam melakukan klasifikasi kelulusan mahasiswa serta mengetahui perbandingan kinerja algoritma *naïve bayes* sebelum dan setelah diterapkan prapemrosesan *random oversampling* dan seleksi fitur.
3. Dapat membantu perguruan tinggi dan fakultas untuk mengetahui variabel yang berpengaruh terhadap kelulusan mahasiswa.

F. Sistematika Penulisan

Laporan penelitian ini ditulis dengan sistematika penulisan secara sistematis yaitu sebagai berikut :

BAB I PENDAHULUAN

Bab ini menjelaskan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II KAJIAN PUSTAKA DAN LANDASAN TEORI

Bab ini menjelaskan tentang kajian hasil penelitian yang telah dilakukan para peneliti lain pada kajian pustaka, sedangkan landasan teori merupakan kajian dari teori-teori para ahli yang dijadikan landasan dalam penelitian.

BAB III METODE PENELITIAN

Bab ini berisi mengenai objek penelitian dan metode yang digunakan dalam penelitian, serta tahapan-tahapan penelitian yang dilakukan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan proses analisa terhadap permasalahan serta solusi yang diberikan, perancangan program berupa perancangan database, implementasi hasil perancangan ke dalam sebuah program serta menjelaskan tentang hasil pengujian dari program yang sudah dibuat.

BAB V PENUTUP

Bab ini berisi kesimpulan yang diperoleh dari pembahasan dalam penelitian serta saran untuk penelitian selanjutnya.

BAB V
PENUTUP

A. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka kesimpulan yang diperoleh dari penelitian ini adalah sebagai berikut.

1. Metode klasifikasi *Naïve Bayes* dapat digunakan untuk mengklasifikasikan kelulusan mahasiswa. Pengujian algoritma *Naïve Bayes* dilakukan untuk mengetahui perbandingan kinerja klasifikasi sebelum dan setelah diterapkan prapemrosesan *Random Oversampling* dan seleksi fitur *Information Gain* dan *Forward Selection*. Berdasarkan hasil pengujian kinerja, diperoleh hasil bahwa kinerja algoritma klasifikasi memiliki kinerja yang baik terhadap *datasets* kelulusan mahasiswa. Algoritma *Naïve Bayes* mengalami peningkatan kinerja setelah dilakukan prapemrosesan *Random Oversampling* dan seleksi fitur. Metode seleksi fitur *Forward Selection* memiliki pengaruh kinerja lebih baik dalam mengklasifikasi data, sehingga dapat mengurangi kompleksitas data dan fitur-fitur yang tidak relevan serta meningkatkan kinerja algoritma klasifikasi *Naïve Bayes*.
2. Akurasi diperoleh pada algoritma *Naïve Bayes* tanpa prapemrosesan *Random Oversampling* dan seleksi fitur dengan metode evaluasi *k-fold cross validation* adalah 81.83%, dengan prapemrosesan *Random Oversampling* menghasilkan akurasi sebesar 83.84%, dengan seleksi fitur *Information Gain* diperoleh akurasi 86.03% dengan 3 fitur terpilih yaitu ip semester 8, ip semester 7, dan ipk, dengan seleksi fitur *Forward Selection* diperoleh akurasi 86.42% dengan 2 fitur terpilih yaitu ip semester 8 dan ipk, sehingga terjadi peningkatan akurasi sebesar 4.2% dari tanpa prapemrosesan ke *Information Gain* dan 4.59% dari tanpa prapemrosesan ke *Forward Selection*. Oleh karena itu, dapat diketahui bahwa tahapan prapemrosesan *Random Oversampling* dan seleksi fitur berpengaruh pada algoritma klasifikasi *Naïve Bayes* dan seleksi fitur *Forward Selection* memiliki pengaruh kinerja lebih baik dengan 2 atribut terpilih dalam

mengklasifikasi data secara tepat, sehingga dapat meningkatkan kinerja algoritma klasifikasi *Naïve Bayes*.

B. Saran

Saran yang diberikan untuk penelitian selanjutnya adalah melakukan prapemrosesan lain seperti *outlier detection* dan prapemrosesan seleksi fitur dengan pendekatan lain selain pendekatan *filter* dan *wrapper* seperti *embedded* untuk mengetahui seleksi fitur yang paling berpengaruh pada klasifikasi. Selain itu diharapkan juga untuk menguji metode klasifikasi selain *Naïve Bayes* untuk mengetahui metode klasifikasi mana yang paling baik kinerjanya.

DAFTAR PUSTAKA

ACCJC/WASC. (2012). *Guide To Evaluating Institutions*. 56. www.g-fras.org

Albarak, M., Alrazgan, M., & Bahsoon, R. (2017). *Identifying and Managing Technical Debt in Database Normalization Using Machine Learning and Trade-off Analysis*. http://arxiv.org/abs/1711.06109

Ali, A., Khairan, A., Tempola, F., & Fuad, A. (2021). Application Of Naïve Bayes to Predict the Potential of Rain in Ternate City. *E3S Web of Conferences*, *328*, 04011. https://doi.org/10.1051/e3sconf/202132804011

Ang, J. C., Mirzal, A., Haron, H., & Hamed, H. N. A. (2016). Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *13*(5), 971–989. https://doi.org/10.1109/TCBB.2015.2478454

Artaye, K. (2015). *International Conference On Information Technology And Business ISSN 2460-7223 IMPLEMENTATION OF NAÏVE BAYES CLASSIFICATION METHOD TO PREDICT GRADUATION TIME OF IBI DARMAJAYA SCHOLAR Z . A . Pagar Alam Street No . 93 Bandar Lampung*. *August*, 284–290.

Ashari, A., Paryudi, I., & Tjoa, A. M. (2013). Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool. *(IJACSA) International Journal of Advanced Computer Science and Applications*, *4*(11), 33–39.

Astuti, F. D. (2018). Seleksi Fitur Forward Selection pada Algoritma Naive Bayes untuk Klasifikasi Benih Gandum. *Jurnal Informasi Interaktif*, *3*(1), 161–166.

BAN-PT. (2015). Akreditasi Institusi Perguruan Tinggi. *Pedoman Penyusunan Borang*.

Bassi, J. S., Dada, E. G., Hamidu, A. A., & Elijah, M. D. (2019). Students Graduation on Time Prediction Model Using Artificial Neural Network. *IOSR Journal of Computer Engineering (IOSR-JCE)*, *21*(3), 28–35. https://doi.org/10.9790/0661-2103012835

Bermejo, P., Gámez, J. A., & Puerta, J. M. (2014). Speeding up incremental wrapper feature subset selection with Naive Bayes classifier. *Knowledge-Based Systems*, *55*, 140–147. https://doi.org/10.1016/j.knosys.2013.10.016

Berrar, D. (2018). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, *1–3*(January 2018), 542–545. https://doi.org/10.1016/B978-0-12-809633-8.20349-X

Bimantoro, D. A., & 'Uyun, S. (2017). *PENGARUH PENGGUNAAN INFORMATION GAINUNTUK SELEKSI FITUR CITRA TANAH DALAM RANGKA MENILAI KESESUAIAN LAHAN PADA TANAMAN CENGKEH*. *2*(1), 42–52.

C.Deisy, B.Subbulakshmi, S.Baskar, D., & Dr.N.Ramaraj. (2007). Efficient Dimensionality Reduction Approaches for Feature Selection. *Proceedings - International Conference on Computational Intelligence and Multimedia Applications, ICCIMA 2007*, *4*, 270–272. https://doi.org/10.1109/ICCIMA.2007.288

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

Fali Oklilas, A., Tasmi, Desy Siswanti, S., Afrina, M., & Setiawan, H. (2019). Attribute selection using information gain and naïve bayes for traffic classification. *Journal of Physics: Conference Series*, *1196*(1). https://doi.org/10.1088/1742-6596/1196/1/012021

Fanani, M. R. (2020). Penggabungan Forward Selection untuk Pemilihan Fitur pada Prediksi Bimbingan Konseling Siswa dengan Menggunakan Algoritma Naive Bayes. *Smart Comp :Jurnalnya Orang Pintar Komputer*, *9*(2), 85–88. https://doi.org/10.30591/smartcomp.v9i2.1924

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1997). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *Journal of Japan Society for Fuzzy Theory and Systems*, *9*(6), 851–860. https://doi.org/10.3156/jfuzzy.9.6_851

Gnana, D. A. A., Singh, Balamurugan, S. A., & Leavline, E. J. (2016). Literature

Review on Feature Selection Methods for High-Dimensional Data. *International Journal of Computer Applications*, *136*(1), 9–17.

Grzymala-Busse, J. W., & Grzymala-Busse, W. J. (2010). Handling Missing Attribute Values. *Data Mining and Knowledge Discovery Handbook*, 33–51. https://doi.org/10.1007/978-0-387-09823-4_3

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. In *Third Edition. 2011. British Library Cataloguing-in-Publication. Morgan Kaufmann*. https://doi.org/10.1109/ICMIRA.2013.45

Harahap, F., Harahap, A. Y. N., Ekadiansyah, E., Sari, R. N., Adawiyah, R., & Harahap, C. B. (2019). Implementation of Naïve Bayes Classification Method for Predicting Purchase. *2018 6th International Conference on Cyber and IT Service Management, CITSM 2018*, *Citsm*, 1–5. https://doi.org/10.1109/CITSM.2018.8674324

Hashemi, H. Z., Parvasideh, P., Larijani, Z. H., & Moradi, F. (2018). Analyze students performance of a national exam using feature selection methods. *2018 8th International Conference on Computer and Knowledge Engineering, ICCKE 2018*, *Iccke*, 7–11. https://doi.org/10.1109/ICCKE.2018.8566671

Henderi, H. (2021). Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *IJIIS: International Journal of Informatics and Information Systems*, *4*(1), 13–20. https://doi.org/10.47738/ijiis.v4i1.73

Heranova, O. (2021). *Synthetic Minority Oversampling Technique pada Averaged One Dependence Estimators untuk Klasifikasi Credit Scoring*. *1*(10), 10–12.

Jain, D., & Singh, V. (2018). Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal*, *19*(3), 179–189. https://doi.org/10.1016/j.eij.2018.03.002

Jurafsky, D., & Martin, J. (2019). Naive bayes and sentiment classification. *Speech and Language Processing*, 1024. https://web.stanford.edu/~jurafsky/slp3/6.pdf

Kalyani, P., & Karnan, D. M. (2010). Attribute Reduction using Forward Selection

and Relative Reduct Algorithm. *International Journal of Computer Applications*, *11*(3), 8–12. https://doi.org/10.5120/1564-1499

Khaire, U. M., & Dhanalakshmi, R. (2019). Stability of feature selection algorithm: A review. *Journal of King Saud University - Computer and Information Sciences*, *xxxx*. https://doi.org/10.1016/j.jksuci.2019.06.012

Khemphila, A., & Boonjing, V. (2011). Heart disease classification using neural network and feature selection. *Proceedings - ICSEng 2011: International Conference on Systems Engineering*, *2007*, 406–409. https://doi.org/10.1109/ICSEng.2011.80

Lasarudin, A., & Purwanto. (2018). KLASIFIKASI PENGADUAN MASYARAKAT MENGGUNAKAN NAIVE BAYES BERBASIS SELEKSI ATRIBUT INFORMATION GAIN. *Jurnal Teknologi Informasi*, *14*(2), 63. http://research.pps.dinus.ac.id/index.php/Cyberku/article/view/65

Lei, C., & Li, K. F. (2015). Academic Performance Predictors. *Proceedings - IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, WAINA 2015*, 577–581. https://doi.org/10.1109/WAINA.2015.114

Mani, K., & Kalpana, P. (2016). An Efficient Feature Selection based on Bayes Theorem, Self Information and Sequential Forward Selection. *International Journal of Information Engineering and Electronic Business*, *8*(6), 46–54. https://doi.org/10.5815/ijieeb.2016.06.06

Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., & Doulamis, N. (2021). Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. *Technologies*, *9*(4), 81. https://doi.org/10.3390/technologies9040081

Muin, A. A. (2016). Penerapan seleksi atribut weights by information gain dan select by weights pada algoritma Naive Bayes untuk prediksi kolektibilitas pembiayaan usaha kecil dan menengah. *Technologia*, *7*(4), 245–249.

Mulyadi, C., & Sukron, . (2020). *Prediction of Timeliness of Graduating with Naïve Bayes Algorithm*. *Icri 2018*, 3043–3050. https://doi.org/10.5220/0009946430433050

Muqorobin, M., Kusrini, K., & Luthfi, E. T. (2019). Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Spp Sekolah. *Jurnal Ilmiah SINUS*, *17*(1), 1. https://doi.org/10.30646/sinus.v17i1.378

Mwadulo, M. W. (2016). A Review on Feature Selection Methods For Classification Tasks. *International Journal of Computer Applications Technology and Research*, *5*(6), 395–402. https://doi.org/10.1109/ICISC.2017.8068746

Naganjaneyulu, S., & Kuppa, M. R. (2013). A novel framework for class imbalance learning using intelligent under-sampling. *Progress in Artificial Intelligence*, *2*(1), 73–84. https://doi.org/10.1007/s13748-012-0038-2

Normawati, D., & Ismi, D. P. (2019). K-Fold Cross Validation for Selection of Cardiovascular Disease Diagnosis Features by Applying Rule-Based Datamining. *Signal and Image Processing Letters*, *1*(2), 23–35. https://doi.org/10.31763/simple.v1i2.3

Nuffic. (2017). *Education system Indonesia*. www.nuffic.nl/en/home/copyright

Nugroho, M. F., & Wibowo, S. (2017). Fitur Seleksi Forward Selection Untuk Menetukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes. *Jurnal Informatika Upgris*, *3*(1), 63–70. https://doi.org/10.26877/jiu.v3i1.1669

Nurjoko, H., & Kurniawan. (2016). *Aplikasi datamining untuk memprediksi tingkat kelulusan mahasiswa menggunakan algoritma apriori di ibi darmajaya bandar lampung*. *02*(01), 79–93.

Panthong, R., & Srivihok, A. (2015). Wrapper Feature Subset Selection for Dimension Reduction Based on Ensemble Learning Algorithm. *Procedia Computer Science*, *72*, 162–169. https://doi.org/10.1016/j.procs.2015.12.117

Patil, T. R., & Sherekar, M. S. S. (2019). Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification. *International Journal of Intelligent Systems and Applications in Engineering*, *7*(2), 88–91. https://doi.org/10.18201/ijisae.2019252786

Prasetiyowati, M. I., Maulidevi, N. U., & Surendro, K. (2021). Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00472-4

Purnana, P., & Supriyanto, D. C. (2013). Deteksi Penyakit Diabetes Type Ii Dengan Naive Bayes Berbasis Particle Swarm Optimization. *Jurnal Teknologi Informasi*, *9*(2), 1414–9999. http://research.pps.dinus.ac.id

Putri, L. R., Mubarok, M. S., & Adiwijaya. (2017). *KLASIFIKASI SENTIMEN ULASAN BUKU BERBAHASA INGGRIS MENGGUNAKAN INFORMATION GAIN DAN NAÏVE BAYES*. *4*(3), 4659–4666.

Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Eeccis*, *7*(1), 59–64. https://doi.org/10.1038/hdy.2009.180

Rozzaqi, A. R. (2015). *Naïve Bayes dan Filtering Feature Selection Information Gain untuk Prediksi Ketepatan Kelulusan Mahasiswa*. 30–41.

Saifudin, A., Ekawati, Yulianti, & Desyani, T. (2020). Forward Selection Technique to Choose the Best Features in Prediction of Student Academic Performance Based on Naïve Bayes. *Journal of Physics: Conference Series*, *1477*(2). https://doi.org/10.1088/1742-6596/1477/3/032007

Seh, A. H. (2019). *A Review on Heart Disease Prediction Using Machine Learning Techniques A Review on Heart Disease Prediction Using Machine Learning Techniques*. *9*(April), 208–224.

Shivali, Joni Birla, G. (2015). Knowledge Discovery in Data-Mining. *International Journal of Engineering Research & Technology (IJERT)*, *3*(10), 1–5. https://www.ijert.org/research/knowledge-discovery-in-data-mining-IJERTCONV3IS10051.pdf

Singh, M., & Gupta, S. (2020). Sentiment Analysis using Naive Bayes Classifier and Information Gain Feature Selection over Twitter. *International Journal of Computer Trends and Technology*, *68*(5), 84–91. https://doi.org/10.14445/22312803/ijctt-v68i5p117

Sugriyono, S., & Siregar, M. U. (2020). Preprocessing kNN algorithm classification

using K-means and distance matrix with students' academic performance dataset. *Jurnal Teknologi Dan Sistem Komputer*, *8*(4), 311–316. https://doi.org/10.14710/jtsiskom.2020.13874

Tangirala, S. (2020). Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm. *International Journal of Advanced Computer Science and Applications*, *11*(2), 612–619. https://doi.org/10.14569/ijacsa.2020.0110277

Ting, S. L., Ip, W. H., & Tsang, A. H. C. (2011). Is Naïve bayes a good classifier for document classification? *International Journal of Software Engineering and Its Applications*, *5*(3), 37–46.

Uyun, S., & Choridah, L. (2018). Feature selection mammogram based on breast cancer mining. *International Journal of Electrical and Computer Engineering*, *8*(1), 60–69. https://doi.org/10.11591/ijece.v8i1.pp60-69

Uyun, S., & Sulistyowati, E. (2020). Feature selection for multiple water quality status: Integrated bootstrapping and SMOTE approach in imbalance classes. *International Journal of Electrical and Computer Engineering*, *10*(4), 4331–4339. https://doi.org/10.11591/ijece.v10i4.pp4331-4339

van Lith, J. W., & Vanschoren, J. (2021). *From Strings to Data Science: a Practical Framework for Automated String Handling*. 1–19. https://arxiv.org/abs/2111.01868v1

Vanaja, S., & Ramesh Kumar, K. (2014). Analysis of Feature Selection Algorithms on Classification: A Survey. *International Journal of Computer Applications*, *96*(17), 29–35. https://doi.org/10.5120/16888-6910

Visalakshi, S., & Radha, V. (2015). A literature review of feature selection techniques and applications: Review of feature selection in data mining. *2014 IEEE International Conference on Computational Intelligence and Computing Research, IEEE ICCIC 2014*, *1997*. https://doi.org/10.1109/ICCIC.2014.7238499

Wah, Y. B., Ibrahim, N., Hamid, H. A., Abdul-Rahman, S., & Fong, S. (2018). Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy. *Pertanika Journal of Science and*

Technology, *26*(1), 329–340.

Wang, Q., Luo, Z., Huang, J., Feng, Y., & Liu, Z. (2017). A Novel Ensemble Method for Imbalanced Data Learning. *Computational Intelligence and Neuroscience*, *2017*, 1–11.

Wang, Y., Li, Y., Song, Y., Rong, X., & Zhang, S. (2017). Improvement of ID3 algorithm based on simplified information entropy and coordination degree. *Algorithms*, *10*(4), 1–18. https://doi.org/10.3390/a10040124

Yunianita, S., Setiani, N., dan Mulyati, S. (2018). Prediksi Ketepatan Masa Studi Mahasiswa dengan Algoritma Pohon Keputusan C45. *SNATi*, 23–29.

Zeniarja, J., Widia, K., & Sani, R. R. (2020). Penerapan Algoritma Naive Bayes dan Forward Selection dalam Pengklasifikasian Status Gizi Stunting pada Puskesmas Pandanaran Semarang. *JOINS (Journal of Information System)*, *5*(1), 1–9. https://doi.org/10.33633/joins.v5i1.2745