

DISERTASI

**PENERAPAN *ATTRIBUTE DEPENDENCY* DAN *ATTRIBUTE ANALYSIS* UNTUK PEMILIHAN FITUR PADA PREDIKSI
KEBERHASILAN AKADEMIK MAHASISWA
MENGUNAKAN METODE KLASIFIKASI**

**(*APPLICATION OF ATTRIBUTE DEPENDENCY AND
ATTRIBUTE ANALYSIS FOR FEATURE SELECTION IN
PREDICTING STUDENT ACADEMIC SUCCESS USING
CLASSIFICATION METHOD*)**

Diajukan untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Doktor dalam Bidang Ilmu Teknik Informatika



DAFID
03013622025009

**PROGRAM STUDI ILMU TEKNIK PROGRAM DOKTOR
FAKULTAS TEKNIK
UNIVERSITAS SRIWIJAYA
NOVEMBER 2023**

LEMBAR PENGESAHAN

PENERAPAN *ATTRIBUTE DEPENDENCY* DAN *ATTRIBUTE ANALYSIS* UNTUK PEMILIHAN FITUR PADA PREDIKSI KEBERHASILAN AKADEMIK MAHASISWA MENGUNAKAN METODE KLASIFIKASI

(*APPLICATION OF ATTRIBUTE DEPENDENCY AND
ATTRIBUTE ANALYSIS FOR FEATURE SELECTION IN
PREDICTING STUDENT ACADEMIC SUCCESS USING
CLASSIFICATION METHOD*)

DISERTASI

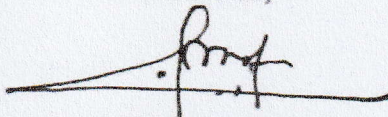
Diajukan untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Doktor dalam Bidang Ilmu Teknik Informatika

Oleh:

DAFID
03013622025009

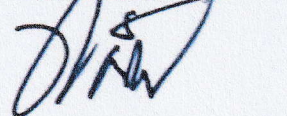
Palembang, 20 November 2023

Promotor,



Dr. Ermatita, M.Kom.
NIP. 196709132006042001

Ko-Promotor



Samsuryadi, S.Si., M.Kom., Ph.D.
NIP. 197102041997021003

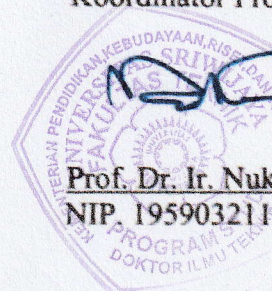

Mengetahui,

Dekan Fakultas Teknik,



Prof. Dr. Eng Ir. H. Joni Arliansyah, MT
NIP. 196706151995121002

Koordinator Program Studi,



Prof. Dr. Ir. Nukman, M.T.
NIP. 195903211987031001

HALAMAN PERSETUJUAN


Karya tulis ilmiah berupa laporan disertasi ini dengan judul "Penerapan *Attribute Dependency* dan *Attribute Analysis* Untuk Pemilihan Fitur Pada Prediksi Keberhasilan Akademik Mahasiswa Menggunakan Metode Klasifikasi" telah dipertahankan di hadapan Tim Penguji Karya Tulis Ilmiah Program Studi Ilmu Teknik Program Doktor Fakultas Teknik Universitas Sriwijaya pada tanggal 15 November 2023.

Palembang, 15 November 2023

Tim Penguji Karya Tulis Ilmiah berupa Laporan Disertasi

Ketua :

Prof. Dr. Ir. Nukman, M.T.
NIP. 195903211987031001

()

Anggota :

1. Dr. Abdiansah, S. Kom., M.Cs
NIP. 198410012009121005

()

2. Dr. Anita Desiani, S.Si., M.Kom
NIP. 197712112003122002


()

3. Dr. Ade Silvia Handayani, S.T., M.T.
NIP. 197609302000032002


()

Mengetahui,

a.n Dekan Fakultas Teknik,


Prof. Dr. Eng. Ir. H. Joni Arliansyah, MT
NIP. 196706151995121002

Koordinator Program Studi,


Prof. Dr. Ir. Nukman, M.T.
NIP. 195903211987031001

KATA PENGANTAR

Alhamdulillah Rabbil Alamin, puji syukur saya panjatkan kehadirat Allah SWT, karena atas berkat rahmat-Nya, saya dapat menyelesaikan disertasi ini. Penulisan disertasi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Doktor pada Fakultas Teknik Universitas Sriwijaya. Saya menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, sangatlah sulit bagi saya untuk menyelesaikan disertasi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Ibu Dr. Ermatita, M.Kom selaku promotor yang telah menyediakan waktu, tenaga, pikiran dan kesabaran untuk mengarahkan saya dalam penyusunan disertasi ini.
2. Bapak Samsuryadi, S.Si., M.Kom., Ph.D yang telah menyediakan waktu, tenaga, pikiran dan kesabaran membimbing saya untuk dapat menyelesaikan disertasi ini.
3. Ketua Program Studi Prof. Dr. Ir. Nukman, M.T. dan para dosen pengajar dari Program Studi Ilmu Teknik S3 BKU Teknik Informatika Universitas Sriwijaya, beserta staf Ibu Yuni yang telah membimbing dan membantu baik selama masa perkuliahan maupun dalam menyelesaikan disertasi.
4. Bapak Alexander Kurniawan dan Bapak James Alexander selaku Ketua dan pihak Yayasan MDP serta Bapak Dr. Johannes Petrus, S.Kom., M.T.I selaku Rektor Universitas MDP yang telah mendukung untuk menyelesaikan pendidikan saya.
5. Rekan-rekan dari Program Studi Ilmu Teknik S3 BKU Teknik Informatika Universitas Sriwijaya yang telah bersama-sama saling memberikan dorongan dan bantuan.
6. Para mentor Universitas MDP yang telah mengarahkan dan memotivasi untuk dapat menyelesaikan disertasi saya.
7. Rekan-rekan Universitas MDP yang telah sangat membantu dalam mengumpulkan data dan bantuan lainnya dalam rangka menyelesaikan disertasi saya.

8. Orang tua saya beserta seluruh keluarga besar saya yang tercinta atas doa dan dukungannya selama ini.
9. Istri saya Veratiwi S.ST, M.Keb tersayang atas kesabaran dan pengertiannya yang sangat besar selama ini dan anak-anak saya tersayang Amirah Adzkiya dan M. Khairan Syauqi yang selalu menjadi penyemangat dalam menyelesaikan disertasi ini.
10. Semua pihak yang telah membantu yang tidak dapat saya tuliskan satu persatu.

Akhir kata, saya berharap Allah SWT berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga disertasi ini membawa manfaat bagi pengembangan ilmu.

Yang menyatakan

Dafid

NIM. 03013622025009

HALAMAN PERNYATAAN INTEGRITAS

Saya yang bertanda tangan dibawah ini :

Nama : Dafid

NIM : 03013622025009

Judul : Penerapan *Attribute Dependency* dan *Attribute Analysis* untuk
Pemilihan Fitur Pada Prediksi Keberhasilan Akademik Mahasiswa
Menggunakan Metode Klasifikasi

Menyatakan bahwa Disertasi saya merupakan hasil karya sendiri didampingi tim promotor dan ko-promotor dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan / plagiat dalam Disertasi ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai aturan yang berlaku.

Demikian, pernyataan ini saya buat dalam keadaan sadar dan tanpa ada paksaan dari siapapun.

Palembang, 20 November 2023

Dafid

NIM. 03013622025009

RINGKASAN

PENERAPAN *ATTRIBUTE DEPENDENCY* DAN *ATTRIBUTE ANALYSIS* UNTUK PEMILIHAN FITUR PADA PREDIKSI KEBERHASILAN AKADEMIK MAHASISWA MENGGUNAKAN METODE KLASIFIKASI

Karya Tulis Ilmiah Berupa Disertasi, tanggal 15 November 2023

Dafid; Dibimbing oleh Promotor Dr. Ermatita, M.Kom, Ko Promotor Samsuryadi S.Si., M.Kom., Ph.D.

Program Studi Ilmu Teknik Program Doktor, Fakultas Teknik Universitas Sriwijaya

Pendidikan memegang peranan yang sangat penting dalam pembangunan sumber daya manusia yang berkualitas di suatu negara. Hal ini salah satunya ditandai dengan keberhasilan mahasiswa dalam menempuh pendidikannya. Kemampuan untuk memprediksi keberhasilan akademik mahasiswa akan menciptakan peluang untuk meningkatkan hasil pendidikan, sehingga perlu diprediksi supaya perguruan tinggi dapat mengambil kebijakan akademik dengan lebih akurat dan efisien. Salah satu metode yang dapat digunakan untuk melakukan prediksi adalah klasifikasi yang ada dalam data mining. Prediksi keberhasilan akademik mahasiswa dilakukan dengan membuat suatu model prediksi yang dihasilkan oleh *classifier* dengan menggunakan dataset pendidikan yang dimiliki oleh perguruan tinggi. Tujuannya adalah untuk mendapatkan informasi/pengetahuan maupun manfaat lain dari dataset tersebut terutama hasil prediksi mahasiswa yang beresiko mengalami kegagalan akademik. Kinerja dari model yang dihasilkan sangat menentukan informasi yang diperoleh oleh perguruan tinggi untuk digunakan lebih lanjut. Dari tiga jenis fitur yang ada dalam suatu dataset, hanya fitur relevan yang harus diseleksi, sedangkan fitur yang *redundant* dan *irrelevant* harus dieliminasi karena mempengaruhi kinerja dari model prediksi yang dihasilkan menjadi tidak optimal.

Teknik yang digunakan untuk menyeleksi dan mengeliminasi fitur tanpa menghilangkan informasi penting yang terkandung di dataset yang berdampak terhadap meningkatnya kinerja dari model prediksi yang dihasilkan adalah *fitur subset selection/feature subset selection* atau yang lebih umum dikenal dengan nama *feature selection*. Dalam prediksi keberhasilan mahasiswa, *filter-based feature selection* merupakan salah satu pendekatan *feature selection* yang umum digunakan bersamaan dengan metode klasifikasi. Pendekatan ini memiliki keunggulan dalam hal *computational cost* yang rendah namun perlu peningkatan pada proses pemilihan fitur untuk mendapatkan fitur terseleksi yang bebas dari fitur *redundant* dan *irrelevant*. Masalah ini dapat diatasi dengan menggunakan teknik *attribute dependency* dan *attribute analysis* yang memiliki kemampuan untuk menentukan fitur tersebut melalui pengukuran ketergantungan dan jenis fitur secara cepat. Hasilnya adalah fitur terseleksi terbaik yang dengan kombinasi metode klasifikasi yang tepat, mampu meningkatkan kinerja model prediksi.

Kata Kunci : Keberhasilan Akademik, Prediksi, Klasifikasi, *Feature Selection*, *Attribute Dependency*, *Attribute Analysis*

SUMMARY

APPLICATION OF ATTRIBUTE DEPENDENCY AND ATTRIBUTE ANALYSIS FOR FEATURE SELECTION IN PREDICTING STUDENT ACADEMIC SUCCESS USING CLASSIFICATION METHOD

Scientific Paper in the form of Dissertation, 15 November 2023

Dafid; supervised by Dr. Ermatita, M.Kom, Co Promotor Samsuryadi S.Si., M.Kom., Ph.D.

Doctoral Program in Engineering Science, Faculty of Engineering Universitas Sriwijaya

Education plays a very important role in developing quality human resources in a country. One way of this is indicated by the success of students in pursuing their studies. The ability to predict student academic success will create opportunities to improve educational outcomes, so it needs to be predicted so that universities can create academic policies more accurately and efficiently. One method that can be used to make predictions is the classification in data mining. Predicting student academic success is done by creating a prediction model produced by a classifier using educational datasets owned by universities. The aim is to obtain information/knowledge and other benefits from the dataset, especially the prediction results of students who are at risk of academic failure. The performance of the resulting model greatly determines the information obtained by universities for further use. Among of the three types of features in a dataset, only relevant features must be selected, while redundant and irrelevant features must be eliminated because they affect the performance of the resulting prediction model to be less than optimal. The technique used to select and eliminate features without losing important information contained in the dataset which has an impact on increasing the performance of the resulting prediction model is feature subset

selection or more commonly known as feature selection. In predicting student success, filter-based feature selection is a feature selection approach that is commonly used in collaboration with classification methods. This approach has the advantage of low computational costs but requires improvements in the feature selection process to obtain selected features that are free from redundant and irrelevant features. This problem can be overcome by using attribute dependency and attribute analysis techniques which have the ability to determine these features through quickly measuring dependencies and feature types. The result is the best selected features which, with the right combination of classification methods, can improve the performance of the prediction model.

Keywords : Academic Success, Prediction, Classification, Feature Selection, Attribute Dependency, Attribute Analysis

DAFTAR ISI

LEMBAR PENGESAHAN	i
HALAMAN PERSETUJUAN.....	ii
KATA PENGANTAR	iii
HALAMAN PERNYATAAN INTEGRITAS.....	v
RINGKASAN	vi
SUMMARY	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
DAFTAR LAMPIRAN.....	xv
BAB 1 PENDAHULUAN	1
1.1 Latar belakang.....	1
1.2 Rumusan Masalah	6
1.3 Tujuan Penelitian	6
1.4 Batasan Masalah.....	7
1.5 Manfaat Penelitian	7
BAB 2 TINJAUAN PUSTAKA	8
2.1 Keberhasilan Akademik.....	8
2.1.1 Ukuran Keberhasilan Akademik	9
2.1.2 Kategori Fitur <i>Predictor</i>	10
2.2 Data Mining	11
2.2.1 Peran Data Mining	11
2.2.2 Algoritma Data Mining	12
2.3 Feature Selection.....	15
2.3.1 <i>Filter-based feature ranking technique (FBFR)</i>	20
2.3.2 <i>Filter-based feature subset technique (FBFS)</i>	22
2.4 <i>Framework</i>	22
2.5 Evaluasi kinerja	22
2.6 <i>Attribute Dependency</i>	23
2.7 <i>Attribute Analysis</i>	25
2.8 Analisis <i>State of The Art</i> dan <i>Research Gap</i>	26
BAB 3 METODOLOGI PENELITIAN.....	33
3.1 Kerangka Penelitian	33
3.2 Dataset.....	39

3.3 <i>Framework</i> Prediksi Keberhasilan Akademik Mahasiswa	44
BAB 4 HASIL YANG DIHARAPKAN.....	46
BAB 5 PEMILIHAN FITUR.....	47
5.1 Ruang dan teknik pencarian fitur	47
5.2 Teknik Pemilihan Fitur DFD	49
5.2.1 Tahap <i>Irrelevance Removal</i>	50
5.2.2 Tahap <i>Redundancy Removal</i>	52
5.2.3 Tahap <i>Relevance Selection</i>	55
BAB 6 HASIL DAN PEMBAHASAN.....	58
6.1 Implementasi Teknik Pemilihan Fitur DFD.....	58
6.2 Penerapan Teknik Pemilihan Fitur DFD Pada Dataset I.....	60
6.2.1 Tahap <i>Irrelevance Removal</i> Pada Dataset I	61
6.2.2 Tahap <i>Redundancy Removal</i> Pada Dataset 1	63
6.2.3 Tahap <i>Relevance Selection</i> Pada Dataset I.....	64
6.3 Penerapan Teknik Pemilihan Fitur DFD Pada Dataset II	69
6.3.1 Tahap <i>Irrelevance Removal</i> Pada Dataset II.....	70
6.3.2 Tahap <i>Redundancy Removal</i> Pada Dataset II.....	70
6.3.3 Tahap <i>Relevance Selection</i> Pada Dataset II	71
6.4 Pemodelan dan Evaluasi.....	76
6.5 Pemodelan dan Evaluasi Pada Dataset I	77
6.5.1 <i>Decision Tree</i>	78
6.5.2 <i>Naive Bayes</i>	79
6.5.3 <i>Artificial Neural Network</i>	79
6.5.4 <i>K-Nearest Neighbor</i>	80
6.5.5 <i>Support Vector Machine</i>	80
6.6 Perbandingan Kinerja Model Teknik DFD Dataset I.....	81
6.7 Perbandingan Kinerja Model Teknik Sebelumnya Dataset I.....	82
6.8 Pemodelan dan Evaluasi Pada Dataset II	86
6.8.1 <i>Decision Tree</i>	86
6.8.2 <i>Naive Bayes</i>	87
6.8.3 <i>Artificial Neural Network</i>	87
6.8.4 <i>K-Nearest Neighbor</i>	88
6.8.5 <i>Support Vector Machine</i>	89
6.9 Perbandingan Kinerja Model Teknik DFD Dataset II	89
6.10 Perbandingan Kinerja Model Teknik Sebelumnya Dataset II.....	91
BAB 7 KESIMPULAN DAN SARAN	94
7.1 Kesimpulan	94
7.2 Saran.....	95
DAFTAR PUSTAKA	96
DAFTAR LAMPIRAN.....	105

DAFTAR GAMBAR

2.1 Model Konseptual Keberhasilan Akademik	8
2.2 Komponen Keberhasilan Akademik dan Ukurannya	9
2.3 Proses Utama <i>Feature Selection</i>	17
2.4 Pendekatan <i>Filter</i>	19
2.5 Pendekatan <i>Wrapper</i>	20
2.6 Pendekatan <i>Embedded</i>	21
3.1 Kerangka Penelitian	34
3.2 Target Level Prediksi dan Tipe Pembelajaran	35
3.3 <i>Data Preparation</i>	38
3.4 <i>Framework</i> Konseptual Prediksi Keberhasilan Akademik Mahasiswa	45
3.5 <i>Framework</i> Prediksi Keberhasilan Akademik Mahasiswa	45
5.1 Teknik Pemilihan Fitur DFD	48
6.1 Kode program <i>python</i> tahap <i>Irrelevance Removal</i>	59
6.2 Kode program <i>python</i> tahap <i>Redundancy Removal</i>	59
6.3 Kode program <i>python</i> tahap <i>Relevance Selection</i>	60
6.4 Kode program <i>python</i> metode klasifikasi	77
6.5 Kode program <i>python</i> evaluasi untuk tiap metode klasifikasi	77
6.6 Perbandingan hasil klasifikasi metode klasifikasi dataset I	82
6.7 Perbandingan jumlah fitur terpilih untuk tiap teknik dataset I	83
6.8 Perbandingan tingkat akurasi tiap teknik pemilihan fitur dataset I.....	85
6.9 Perbandingan hasil klasifikasi metode klasifikasi dataset II	90
6.10 Perbandingan jumlah fitur terpilih untuk tiap teknik dataset II	92
6.11 Perbandingan tingkat akurasi tiap teknik pemilihan fitur dataset II	93

DAFTAR TABEL

2.1 Kategori Fitur <i>Predictor</i>	10
2.2 Representasi Data Set	16
2.3 Evaluasi Penelitian Teknik <i>Feature Selection</i>	26
2.4 Evaluasi Teknik <i>Filter Based Feature Selection</i>	30
2.5 Evaluasi Penelitian Teknik <i>Attribute Dependency</i> dan <i>Attribute Analysis</i>	31
3.1 Dataset <i>Students' Performance</i>	39
3.2 Dataset akademikMDP	43
5.1 Ruang dan teknik pencarian fitur	48
6.1 Penentuan jenis fitur dataset (D) pada dataset I	60
6.2 Kategori fitur <i>predictor</i> pada dataset I	61
6.3 Penentuan fitur <i>irrelevant</i> (IR) pada dataset I	62
6.4 Dataset hasil eliminasi fitur <i>irrelevant</i> IR (D1) pada dataset I	63
6.5 Penentuan fitur <i>redundant</i> (RD) pada dataset I.....	64
6.6 Dataset hasil eliminasi fitur <i>redundant</i> RD (D2) pada dataset I	64
6.7 Penentuan fitur relevan (RL) pengukuran pertama pada dataset I	65
6.8 Penentuan fitur relevan (RL) pengukuran kedua pada dataset I	66
6.9 Dataset hasil pemilihan fitur relevan RL (D3) pada dataset I	68
6.10 Penentuan jenis fitur dataset (D) pada dataset II	69
6.11 Kategori fitur <i>predictor</i> pada dataset II	69
6.12 Penentuan fitur <i>irrelevant</i> (IR) pada dataset II	70
6.13 Penentuan fitur <i>redundant</i> (RD) pada dataset II	71
6.14 Penentuan fitur relevan (RL) pengukuran pertama pada dataset II.....	71
6.15 Penentuan fitur relevan (RL) pengukuran kedua pada dataset II.....	73
6.16 Dataset hasil pemilihan fitur relevan RL (D3) pada dataset II.....	76
6.17 <i>Confusion matrix</i> klasifikasi <i>Decision Tree</i>	78
6.18 <i>Confusion matrix</i> klasifikasi <i>Naive Bayes</i>	79
6.19 <i>Confusion matrix</i> klasifikasi <i>Artificial Neural Network</i>	79
6.20 <i>Confusion matrix</i> klasifikasi <i>K-Nearest Neighbor</i>	80
6.21 <i>Confusion matrix</i> klasifikasi <i>Support Vector Machine</i>	81
6.22 Perbandingan hasil klasifikasi metode klasifikasi dataset I	82

6.23 Perbandingan jumlah fitur terpilih untuk tiap teknik dataset I.....	83
6.24 Perbandingan tingkat akurasi tiap teknik pemilihan fitur dataset I	84
6.25 <i>Confusion matrix</i> klasifikasi <i>Decision Tree</i>	86
6.26 <i>Confusion matrix</i> klasifikasi <i>Naive Bayes</i>	87
6.27 <i>Confusion matrix</i> klasifikasi <i>Artificial Neural Network</i>	88
6.28 <i>Confusion matrix</i> klasifikasi <i>K-Nearest Neighbor</i>	88
6.29 <i>Confusion matrix</i> klasifikasi <i>Support Vector Machine</i>	89
6.30 Perbandingan hasil klasifikasi metode klasifikasi dataset II	90
6.31 Perbandingan jumlah fitur terpilih untuk tiap teknik dataset II	91
6.32 Perbandingan tingkat akurasi tiap teknik pemilihan fitur dataset II	93

DAFTAR LAMPIRAN

1. <i>Output</i> program klasifikasi menggunakan teknik DFD dataset I.....	105
2. <i>Output</i> program klasifikasi menggunakan teknik DFD dataset II	112
3. Seminar Karya Ilmiah Internasional I	120
4. Seminar Karya Ilmiah Internasional II	125
5. Publikasi Jurnal Internasional Bereputasi	131
6. Surat Keterangan Perbaikan Ujian Disertasi	141

BAB 1

PENDAHULUAN

1.1 Latar belakang

Pendidikan memegang peranan yang sangat penting dalam pembangunan sumber daya manusia yang berkualitas di suatu negara. Hal ini sesuai dengan tujuan pendidikan yang menjadi tujuan pendidikan nasional yang tertuang dalam Pasal 3 UU No 20 Tahun 2003 yaitu “*Pendidikan nasional berfungsi mengembangkan kemampuan dan membentuk watak serta peradaban bangsa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa, bertujuan untuk berkembangnya potensi peserta didik agar menjadi manusia yang beriman dan bertakwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, cakap, kreatif, mandiri dan menjadi warga Negara yang demokratis serta bertanggung jawab*”. Untuk mencapai tujuan tersebut diperlukan suatu penyelenggara pendidikan yang mampu mendorong mahasiswa meraih keberhasilan akademik melalui suatu wadah yang bernama perguruan tinggi.

Dalam penyelenggaraannya, perguruan tinggi menginginkan setiap mahasiswa yang menempuh pendidikan mendapatkan keberhasilan akademik. Namun faktanya angka keberhasilan akademik mahasiswa masih perlu ditingkatkan sehingga perlu upaya lebih lanjut untuk mengurangi angka kegagalan akademik mahasiswa (DirjenDikti, 2020; DirjenDiktiRistek 2021; Kemenristekdikti, 2017, 2018, 2019). Kemampuan untuk memprediksi keberhasilan akademik mahasiswa akan menciptakan peluang bagi perguruan tinggi untuk meningkatkan hasil pendidikan melalui pembuatan kebijakan akademik yang lebih akurat dan efisien (Hellas et al., 2018; Shahiri et al., 2015). Prediksi keberhasilan akademik mahasiswa dilakukan dengan membuat suatu model prediksi menggunakan metode klasifikasi dan dataset pendidikan yang dimiliki oleh perguruan tinggi. Menurut Larose (2005) penggunaan metode tersebut didasarkan pada hasil prediksi yang berupa data kategorik/nominal.

Menurut Alyahyan dan Düşteğör (2020) metode klasifikasi yang digunakan peneliti saat ini untuk memprediksi keberhasilan akademik mahasiswa yaitu *Decision Tree* (44%), *Naive Bayes* (19%), *Artificial Neural Network* (10%), *K-*

Nearest Neighbor (5%) dan *Support Vector Machine* (3%). Dari hasil penelitian menunjukkan bahwa tingkat akurasi menjadi pertimbangan utama dalam membuat model prediksi untuk dapat ditingkatkan lagi (Alyahyan dan Düşteğör, 2020; Namoun dan Alshantiti, 2021; Shahiri et al., 2015). Sehubungan dengan itu, telah ada usaha oleh peneliti untuk meningkatkan akurasi dari model prediksi. Hal ini didasari oleh fakta bahwa fitur yang ada dalam suatu dataset yang digunakan dalam pembuatan model prediksi tidak seluruhnya memuat informasi penting (relevan). Selain fitur yang relevan, dataset tersebut masih memiliki fitur yang *redundant* dan *irrelevant* (Beniwal dan Arora, 2012; Gajwani dan Chakraborty, 2021). Menurut Manikandan dan Abirami (2021) fitur ini mempengaruhi kinerja dari model prediksi yang dihasilkan menjadi tidak optimal, yang mana kinerjanya sangat bergantung dari seleksi fitur yang paling relevan dari fitur-fitur yang ada di dataset (Anukrishna dan Paul, 2017; Srivastava et al., 2013; Zaffar et al., 2018; Zheng, 2018). Untuk menyeleksi fitur relevan dan meminimalisir fitur *redundant* dan *irrelevant* di dataset digunakan teknik pemilihan fitur (*atribut subset selection/feature subset selection*) atau yang lebih umum dikenal dengan nama *feature selection* (Aggarwal, 2016; Larose, 2005). Teknik ini terbukti dapat meningkatkan nilai akurasi suatu model prediksi ((Febro, 2019; Hussain et al., 2018; Khasanah dan Harwati, 2017; Priyasadie dan Isa, 2021; Sökkhey dan Okazaki, 2020; Zaffar, Hashmani, dan Savita, 2018; Zaffar, Hashmani, Savita, et al., 2018).

Feature selection memiliki tiga pendekatan yaitu *filter*, *wrapper* dan *embedded* (Aggarwal, 2016). Dari penelitian yang telah dilakukan sebelumnya oleh Borkar dan Rajeswari (2014); Punlumjeak dan Rachburee (2015); Sökkhey dan Okazaki (2020) menunjukkan bahwa secara umum tiap pendekatan mampu menyeleksi fitur yang relevan walaupun belum secara optimal sehingga berdampak terhadap akurasi model prediksi yang dihasilkan. Namun masing-masing pendekatan memiliki kelebihan dan kekurangan. Pendekatan *filter*, proses yang dilakukan tidak menggunakan *classifier*, sehingga memiliki kelebihan dari waktu komputasi yang cepat, tapi memiliki kekurangan dari tingkat akurasi yang bisa lebih rendah dari pendekatan *wrapper* dan *embedded*. Sebaliknya pendekatan *wrapper* dan *embedded*, proses yang dilakukan menggunakan *classifier* yang

berulang untuk mendapatkan fitur terpilih, sehingga memiliki kelebihan dari tingkat akurasi yang dihasilkan bisa lebih tinggi dari pendekatan *filter*, namun memiliki kekurangan pada waktu komputasi yang lama, sehingga tidak praktis digunakan pada dataset yang berdimensi tinggi (Alyahyan dan Düşteğör, 2020; Xue et al., 2016). Dataset yang dimiliki oleh perguruan tinggi merupakan dataset yang berdimensi tinggi, sehingga sebagian besar peneliti menggunakan pendekatan *filter* untuk pembuatan modelnya.

Teknik *feature selection* pendekatan *filter* yang digunakan dalam bidang pendidikan dikelompokkan lagi menjadi *filter-based feature ranking (FBFR)* dan *filter-based feature subset (FBFS)* dimana kedua teknik tersebut dibedakan dari jumlah fitur dan teknik evaluasi fitur yang digunakan. Teknik *feature selection* pendekatan filter yang digunakan oleh para peneliti dalam dunia pendidikan untuk kategori FBFR yaitu *ChiSquared, Information Gain, Gain Ratio, Symmetrical Uncertainty, Relief, Relief A, Relief F*, sedangkan untuk kategori FBFS yaitu *Correlation Based Feature Selection* dan *Minimum Redundancy Maximum Relevance* (Ali et al., 2019; Amrieh et al., 2016; Anuradha dan Velmurugan, 2016; Arif et al., 2021; Enaro dan Chakraborty, 2020; Febro, 2019; Hussain et al., 2018; Jalota dan Agrawal, 2021; Khasanah dan Harwati, 2017; Nuankaew dan Thongkam, 2020; Priyasadie dan Isa, 2021; Punlumjeak dan Rachburee, 2015; Rachburee dan Punlumjeak, 2015; Rahman et al., 2018; Sökkhey dan Okazaki, 2020; Zaffar et al., 2018).

Masing-masing teknik memiliki kelebihan dan kekurangan tersendiri dalam pemilihan fitur dimana secara umum untuk kategori *filter-based feature ranking technique (FBFR)* kelebihanannya adalah *computational cost* yang rendah namun kekurangannya adalah belum menangani fitur *redundant* dan belum optimal menangani fitur yang *irrelevant* (Ahmed et al., 2017; Alhaj et al., 2016; Ambusaidi et al., 2016; Amrieh et al., 2016; Azhagusundari dan Thanamani, 2013; Karegowda et al., 2010; Manikandan dan Abirami, 2021). Hal ini dipengaruhi teknik yang digunakan dalam menghasilkan fitur terpilih hanya berdasarkan ranking dari *importance score* setiap fitur. Sebaliknya kategori *filter-based feature subset technique (FBFS)* memiliki kelebihan telah ada usaha untuk menangani fitur *redundant* dengan cara memfilter secara keterhubungan dan yang menjadi

kekurangannya adalah masih belum optimal untuk menangani fitur *redundant* dan *irrelevant* serta *evaluation measure* yang lama dan *computational cost* yang tinggi. Hal ini dipengaruhi teknik yang digunakan dalam menghasilkan fitur terpilih dilakukan dengan melakukan iterasi terhadap subset fitur berdasarkan *feature subset evaluator* tertentu (Febro, 2019; Shiva Darshan dan Jaidhar, 2018; Sokkhey dan Okazaki, 2020; Venkatesh dan Anuradha, 2019; Xue et al., 2016).

Masih adanya fitur *redundant* dan *irrelevant* yang muncul di kedua teknik (FBFS dan FBFR) tersebut menunjukkan bahwa kedua teknik tersebut masih perlu diperbaiki pada metode seleksi fiturnya. Selain itu fitur tersebut juga berdampak terhadap tingkat akurasi dari model prediksi yang dibuat menjadi tidak optimal, sehingga fitur tersebut harus dieliminasi supaya masalahnya tidak terjadi. Adapun yang menjadi faktor penyebab masih adanya kedua fitur tersebut adalah adanya relasi diantara fitur yang akan diseleksi yang dapat muncul baik secara tunggal ataupun berkelompok. Untuk itu maka diperlukan suatu teknik yang dapat menentukan fitur mana saja pada suatu dataset yang memiliki relasi dengan fitur lain. Sementara itu dalam model relasional terdapat suatu teknik yang bernama *Attribute Dependency* dimana teknik ini berkaitan erat dengan fitur yang ada pada suatu dataset. *Attribute dependency* (ketergantungan atribut) adalah metode/teknik yang dipergunakan untuk melihat ketergantungan dari suatu fitur terhadap fitur lain (Connolly dan Begg, 2015; Elmasri dan Navathe, 2016; Silberschatz et al., 2010). Kemampuan teknik ini lebih lanjut dapat digunakan untuk menentukan keberadaan relasi antar fitur pada suatu dataset sehingga teknik ini berpeluang dan sesuai untuk mengatasi keterbatasan yang ada pada kedua teknik FBFS dan FBFR. Hal ini didukung oleh hasil penelitian yang dilakukan oleh Berens et al. (2022); Vassiliou (2015) yang menunjukkan bahwa metode ini efektif dalam menentukan keterhubungan dalam sekumpulan atribut untuk menentukan redundansi data. Sifat keterhubungan yang ditunjukkan oleh *attribute dependency* juga sangat membantu dalam menentukan interaksi antar sekumpulan atribut dimana hal ini dapat digunakan untuk mengetahui mana kelompok atribut yang merupakan atribut *complimentarity* (Di et al., 2010; Fan et al., 2008; Kupka dan Tomanová, 2012; Roblot dan Link, 2018). Selain itu dalam proses pencarian keterhubungan data, *attribute dependency* tidak memerlukan pemeriksaan ke seluruh *tuple* sehingga

mampu menjadikannya sebagai *fast evaluation measure* (Fan et al., 2008; Li et al., 2014; Vassiliou, 2015). Lebih lanjut, penelitian lain oleh Belohlavek dan Vychodil (2008) menunjukkan bahwa secara efektif *attribute dependency* dapat digunakan untuk korelasi linear maupun yang non linear. *Attribute dependency* dalam penerapannya dapat melibatkan aksioma *Armstrong* sehingga dapat menemukan keterhubungan data dengan lebih cepat (Belohlavek et al., 2016; Roblot dan Link, 2018; Yamaguchi, 2009).

Relasi antar fitur pada suatu dataset, selain dapat ditentukan dengan mengukur ketergantungan suatu fitur juga dapat diukur dengan menentukan jenis fitur tersebut berdasarkan nilainya. Fitur yang termasuk kategori unik atau yang merupakan turunan dari fitur lain akan menjadikan suatu fitur itu menjadi fitur yang *redundant* atau *irrelevant*. Sementara itu dalam model relasional terdapat suatu teknik yang dapat menentukan kategori suatu fitur yang bernama *Attribute Analysis*. Dalam penggunaannya teknik ini pun berpeluang dan sesuai untuk mengatasi keterbatasan yang ada pada teknik FBFS dan FBFR diatas untuk melengkapi teknik *Attribute Dependency*. *Attribute analysis* merupakan suatu metode untuk menganalisis suatu atribut berdasarkan pengelompokkan tertentu berdasarkan nilai yang dimiliki, yaitu keunikan nilai dan kemampuan untuk menurunkan atribut lain dari suatu atribut (Connolly dan Begg, 2015; Elmasri dan Navathe, 2016; Silberschatz et al., 2010). Penelitian yang dilakukan oleh Mathkour (2009); Tourir dan Mathkour (2008) menunjukkan teknik ini efektif dalam menentukan keunikan nilai suatu atribut (*uniqueness*), *multivalued attribute* dan atribut turunan yang menggunakan pendekatan *database* melalui normalisasi model relasional. Hasil penelitian lain juga menunjukan teknik ini efektif dalam penentuan keunikan atribut (Caldarone dan Williams, 2010; Martinez-torres dan Toral, 2019; Olmedilla et al., 2019). Keunikan suatu nilai atribut (*uniqueness*) dan *derived attribute* selanjutnya dapat digunakan untuk mengatasi belum optimalnya penanganan fitur *redundant* dan *irrelevant*.

Dengan demikian untuk mengatasi belum optimalnya penanganan fitur *redundant* dan *irrelevant* dalam proses pemilihan fitur maka kombinasi dari teknik ketergantungan atribut (*Attribute Dependency*) dan analisa atribut (*Attribute*

Analysis) dapat digunakan sehingga pemilihan fitur bisa lebih optimal yang mampu meningkatkan kinerja dari model prediksi keberhasilan akademik mahasiswa.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan diatas maka dapat dirumuskan masalah yang akan dibahas pada penelitian ini adalah:

1. Bagaimana mengoptimalkan pemilihan fitur untuk prediksi keberhasilan akademik mahasiswa?
2. Bagaimana teknik *Attribute Dependency* dan *Attribute Analysis* dapat digunakan untuk pemilihan fitur?
3. Bagaimana pengaruh penerapan teknik *Attribute Dependency* dan *Attribute Analysis* untuk meningkatkan kinerja model prediksi keberhasilan akademik mahasiswa?
4. Apa metode klasifikasi yang menghasilkan peningkatan kinerja tertinggi dalam memprediksi keberhasilan akademik mahasiswa?
5. Bagaimana pengaruh penerapan teknik *Attribute Dependency* dan *Attribute Analysis* dibandingkan dengan teknik *feature selection* sebelumnya?

1.3 Tujuan Penelitian

Tujuan yang akan dicapai dalam penelitian ini adalah:

1. Mengoptimalkan pemilihan fitur dengan menentukan urutan proses pemilihan fitur.
2. Menentukan kategori suatu fitur yang termasuk fitur yang *irrelevant*, *redundant* atau relevan melalui pengukuran nilai dan ketergantungan fitur.
3. Mendapatkan fitur terpilih yang bebas dari fitur *irrelevant* dan *redundant* yang dapat meningkatkan kinerja model prediksi keberhasilan akademik mahasiswa.
4. Menentukan metode klasifikasi terbaik yang menghasilkan model prediksi keberhasilan akademik mahasiswa dengan kinerja tertinggi.
5. Menentukan besarnya peningkatan kinerja model prediksi yang diusulkan dengan model prediksi sebelumnya.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Penelitian ini dilakukan pada mahasiswa perguruan tinggi
2. Pendekatan pemilihan fitur yang digunakan adalah pendekatan *filter-based feature ranking (FBFR)*
3. Dataset yang digunakan adalah dataset *academic performance* yang merupakan dataset publik dan dataset akademikMDP yang merupakan dataset *private*
4. Metode klasifikasi yang digunakan adalah *Decision Tree, Neural Network, Naive Bayes, K-Nearest Neighbor* dan *Support Vector Machine*
5. Evaluasi kinerja model prediksi yang digunakan adalah *accuracy, precision, recall dan F-measure*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Mengefisienkan ruang pencarian (*search space*) pemilihan fitur sehingga *computational cost* menjadi lebih rendah.
2. Memudahkan dalam mengeliminasi dan memilih fitur untuk pembuatan model prediksi.
3. Meningkatkan kualitas perguruan tinggi dalam menghasilkan lulusan melalui pengambilan kebijakan atau pengembangan strategi pembelajaran berdasarkan akurasi dari model prediksi yang dihasilkan.
4. Membantu perguruan tinggi mendapatkan model prediksi terbaik yang memberikan informasi yang lebih akurat untuk pengambilan keputusan.
5. Meningkatkan kualitas informasi model prediksi yang digunakan oleh perguruan tinggi untuk pengambilan keputusan.

DAFTAR PUSTAKA

- Abusamra. 2013. A comparative study of feature selection and classification methods for gene expression data of glioma. *Procedia Computer Science*, 23, 5–14. <https://doi.org/10.1016/j.procs.2013.10.003>
- Adekitan, 2018. Data mining approach to predicting the performance of first year student in a university using the admission requirements. *Education and Information Technologies*, 2, 11-16. <https://doi.org/10.1007/s10639-018-9839-7> Data
- Aggarwal, 2016. *Data Mining: The Textbook* (Issue April). Springer International Publishing Switzerland.5-24
- Ahmad, Ismail, dan Aziz. 2015. The Prediction of Students' Academic Performance Using Classification Data Mining Techniques. *Applied Mathematical Sciences*, 9, 6415–6426. <https://doi.org/10.12988/ams.2015.53289>
- Ahmed, Riyad dan Shamsudeen. 2017. Information gain based feature selection for intrusion detection system. *International Journal of Scientific & Engineering Research*. 8, 294–298.
- Ajibade, Ahmad, dan Shamsuddin. 2019. An Heuristic Feature Selection Algorithm to Evaluate Academic Performance of Students. *IEEE 10th Control and System Graduate Research Colloquium (ICSGRC)*, August, 110–114.
- Al-barrak, Al-razgan, M. 2016. Predicting Students Final GPA Using Decision Trees : A Case Study. *International Journal of Information and Education Technology*, 6, 528–533. <https://doi.org/10.7763/IJiet.2016.V6.745>
- Alhaj, T. A., Siraj, M., Zainal, A., Elshoush, H. T., dan Elhaj, F. 2016. Feature Selection Using Information Gain for Improved Structural-Based Alert Correlation. *PLoS ONE*. 1–18. <https://doi.org/10.1371/journal.pone.0166017>
- Ali, U., Arif, K. S., dan Qamar, U. 2019. A hybrid scheme for feature selection of high dimensional educational data. *International Conference on Communication Technologies (ComTech)*, 71–75. <https://doi.org/10.1109/Comtech.2019.8737829>
- Almarabeh, H. 2017. Analysis of Students ' Performance by Using Different Data Mining Classifiers. *I.J. Modern Education and Computer Science*, 8(August), 8–15. <https://doi.org/10.5815/ijmecs.2017.08.02>
- Aluko, R. O. 2018. Towards reliable prediction of academic performance of architecture students using data mining techniques. *Journal of Engineering, Design and Technology*. 15, 38-42. <https://doi.org/10.1108/JEDT-08-2017-0081>
- Alyahyan, E., dan Düşteğör, D. 2020. Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1). <https://doi.org/10.1186/s41239-020-0177-7>
- Ambusaidi, M. A., He, X., Nanda, P., dan Tan, Z. 2016. Building an intrusion detection system using a filter-based feature selection algorithm. *IEEE Transactions On Computers*. 65, 1–13. <https://doi.org/10.1109/TC.2016.2519914>
- Amrieh, E. A., Hamtini, T., dan Aljarah, I. 2016. Mining educational data to predict students' academic performance. *International Journal of Database Theory and*

- Application, 9(8), 119–136. https://doi.org/10.1007/978-3-319-21024-7_28
- Anukrishna, P. R., dan Paul, V. 2017. A review on feature selection for high dimensional data. *Proceedings of the International Conference on Inventive Systems and Control, ICISC 2017*, 5(6), 395–402. <https://doi.org/10.1109/ICISC.2017.8068746>
- Anuradha, C., dan Velmurugan, T. 2015. A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance. *Indian Journal of Science and Technology*. 8. <https://doi.org/10.17485/ijst/2015/v8i15/74555>
- Anuradha, C., dan Velmurugan, T. 2016. Performance Evaluation of Feature Selection Algorithms in Educational Data Mining. *International Journal of Data Mining Techniques and Applications*, 5(2), 131–139. <https://doi.org/10.20894/ijdmta.102.005.002.007>
- Arif, M. A., Jahan, A., Mau, M. I., dan Tummarzia, R. 2021. An Improved Prediction System of Students' Performance Using Classification model and Feature Selection Algorithm. *International Journal of Advances in Soft Computing and Its Applications*, 13(1), 162–177.
- Arsad, P. M., Buniyamin, N., dan Manan, J. L. A. 2013. A neural network students' performance prediction model (NNSPPM). *IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, November. <https://doi.org/10.1109/ICSIMA.2013.6717966>
- Asif, R., dan Merceron, A. 2015. Predicting Student Academic Performance at Degree Level : A Case Study. *International Journal of Intelligent Systems Technologies and Applications*. 2. 49-61 <https://doi.org/10.5815/ijisa.2015.01.05>
- Asif, R., Merceron, A., Ali, S. A., dan Haider, N. G. 2017. Analyzing undergraduate students' performance using educational data mining. *Computers dan Education*. 3. 15-20. <https://doi.org/10.1016/j.compedu.2017.05.007>
- Azhagusundari, B., dan Thanamani, A. S. 2013. Feature Selection Based on Information Gain. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*. 2(2), 18–21.
- Badan Akreditasi Perguruan Tinggi. 2019. Akreditasi Perguruan Tinggi (Issue November). http://files/284/Lampiran-02-PerBAN-PT-3-2019-Kriteria-dan-Prosedur-IAPT-3_0.pdf
- Belohlavek, R., Cordero, P., Enciso, M., dan Mora, Á. 2016. Automated prover for attribute dependencies in data with grades. *International Journal of Approximate Reasoning*, 70, 51–67. <https://doi.org/10.1016/j.ijar.2015.12.007>
- Belohvavek, R., dan Vychodil, Vi. 2008. Basic Algorithm For Attribute Implications And Functional Dependencies In Graded Setting. *International Journal of Foundations of Computer Science*. 19(2), 297–317.
- Beniwal, S., dan Arora, J. 2012. Classification and Feature Selection Techniques in Data Mining. *International Journal of Engineering Research dan Technology (IJERT)*, 1(6), 1–6.
- Berens, M., Biskup, J., dan Preuß, M. 2022. Uniform probabilistic generation of relation instances satisfying a functional dependency. *Information Systems*, 11, 20-30. <https://doi.org/10.1016/j.is.2021.101848>
- Borkar, S., dan Rajeswari, K. 2014. Attributes Selection for Predicting Studentsdanapos; Academic Performance using Education Data Mining and

- Artificial Neural Network. *International Journal of Computer Applications*, 86(10), 25–29. <https://doi.org/10.5120/15022-3310>
- Bunkar, K., Singh, U. K., Pandya, B., dan Bunkar, R. 2012. Data mining: Prediction for performance improvement of graduate students using classification. *IFIP International Conference on Wireless and Optical Communications Networks, WOCN*, 3–7. <https://doi.org/10.1109/WOCN.2012.6335530>
- Caldarone, C. A., dan Williams, W. G. 2010. The Congenital Heart Surgeons Society Datacenter : Unique Attributes as a Research Organization. *YPCSU*, 13(1), 71–75. <https://doi.org/10.1053/j.pcsu.2010.01.007>
- Chen, Z., Chen, Q., Zhang, Y., Zhou, L., Jiang, J., Wu, C., dan Huang, Z. 2021. Clustering-based feature subset selection with analysis on the redundancy–complementarity dimension. *Computer Communications*, 168(January), 65–74. <https://doi.org/10.1016/j.comcom.2021.01.005>
- Connolly, T., dan Begg, C. 2015. *Database Systems A Practical Approach to Design, Implementation, and Management*. Pearson Education.
- Dash, M., dan Liu, H. 1997. Feature selection for classification. *Intelligent Data Analysis*, 1(3), 131–156. <https://doi.org/10.3233/IDA-1997-1302>
- Dash, Manoranjan, dan Liu, H. 2003. Consistency-based search in feature selection. *Artificial Intelligence*, 151(1–2), 155–176. [https://doi.org/10.1016/S0004-3702\(03\)00079-1](https://doi.org/10.1016/S0004-3702(03)00079-1)
- Di, F., Loia, V., dan Sessa, S. 2010. Fuzzy transforms method and attribute dependency in data analysis. *Information Sciences*, 180(4), 493–505. <https://doi.org/10.1016/j.ins.2009.10.012>
- Elakia, Gayathri, Aarthi, dan Naren J. 2014. Application of Data Mining in Educational Database for Predicting Behavioural Patterns of the Students. *International Journal of Computer Science and Information Technologies*. 5. 4649-4652
- Elmasri, R., dan Navathe, S. B. 2016. *Fundamentals of Database Systems (7th ed.)*. Pearson.
- Enaro, A. O., dan Chakraborty, S. 2020. Feature selection algorithms for predicting students academic performance using data mining techniques. *International Journal of Scientific and Technology Research*, 9(4), 3622–3626.
- Fan, W., Geerts, F., Jia, X., dan Kementsietsidis, A. 2008. Conditional functional dependencies for capturing data inconsistencies. *ACM Transactions on Database Systems*, 5. 1-10
- Febro, J. D. 2019. Utilizing feature selection in identifying predicting factors of student retention. *International Journal of Advanced Computer Science and Applications*, 10(9), 269–274. <https://doi.org/10.14569/ijacsa.2019.0100934>
- Francis, B. K., dan Babu, S. S. 2019. Predicting Academic Performance of Students Using a Hybrid Data Mining Approach. *Journal of Medical Systems*, 4. 1-15. <https://doi.org/10.1007/s10916-019-1295-4>
- Gajwani, J., dan Chakraborty, P. 2021. Students' Performance Prediction Using Feature Selection and Supervised Machine Learning Algorithms. In *Advances in Intelligent Systems and Computing*. 1165, 347–354. https://doi.org/10.1007/978-981-15-5113-0_25
- Garg, R. 2018. Predicting Student Performance Of Different Regions Of Punjab Using Classification Techniques. *International Journal of Advanced Research in Computer Science*. 9(1), 236–240.

- <https://doi.org/http://dx.doi.org/10.26483/ijarcs.v9i1.5234>
- Ghotra, B., McIntosh, S., dan Hassan, A. E. 2017. A large-scale study of the impact of feature selection techniques on defect classification models. *IEEE International Working Conference on Mining Software Repositories*, 5(1), 146–157. <https://doi.org/10.1109/MSR.2017.18>
- Gray, G., McGuinness, C., dan Owende, P. 2014. An application of classification models to predict learner progression in tertiary education. *Souvenir of the 2014 IEEE International Advance Computing Conference (IACC)*, 12, 549–554. <https://doi.org/10.1109/IAAdCC.2014.6779384>
- Hämäläinen, W., dan Vinni, M. 2006. Comparison of machine learning methods for intelligent tutoring systems. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4, 525–534. https://doi.org/10.1007/11774303_52
- Hamoud, A. K., Hashim, A. S., dan Awadh, W. A. 2018. Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis. *International Journal of Interactive Multimedia and Artificial Intelligence*. 5(2). 26-31. <https://doi.org/10.9781/ijimai.2018.02.004>
- Han, J., Kamber, M., dan Pei, J. 2012. *Data Mining: Concepts and Techniques*. In *Data Mining: Concepts and Techniques*. Morgan Kaufmann. <https://doi.org/10.1016/C2009-0-61819-5>
- Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C., dan Liao, S. N. 2018. Predicting academic performance: A systematic literature review. *Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE)*, 10, 175–199. <https://doi.org/10.1145/3293881.3295783>
- Hussain, S., Dahan, N. A., Ba-Alwib, F. M., dan Ribata, N. 2018. Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, 9(2), 447–459. <https://doi.org/10.11591/ijeecs.v9.i2.pp447-459>
- Imran, M., Latif, S., Mehmood, D., dan Shah, M. S. 2019. Student academic performance prediction using supervised learning techniques. *International Journal of Emerging Technologies in Learning*, 14(14), 92–104. <https://doi.org/10.3991/ijet.v14i14.10310>
- Jalota, C., dan Agrawal, R. 2021. Feature Selection Algorithms and Student Academic Performance: A Study. *International Conference on Innovative Computing and Communications*. 5.317-328. <https://doi.org/10.1007/978-981-15-5113-0>
- Jishan, S. T., Rashu, R. I., Haque, N., dan Rahman, R. M. 2015. Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. *Decision Analytics*, 2(1), 1–25. <https://doi.org/10.1186/s40165-014-0010-2>
- Karegowda, A. G., Manjunath, A. S., dan Jayaram, M. A. 2010. Comparative Study of Attribute Selection Using Gain Ratio and Correlation Based Feature Selection. *International Journal of Information Technology and Knowledge Management*, 2(2), 271–277.
- Kemendikbud. 2020. *Statistik Pendidikan Tinggi (Higer Education Statistic) 2020*. PDDikti Kemendikbud, 5, 81–85. <https://pddikti.kemdikbud.go.id/publikasi>
- Kemenristekdikti. 2017. *Statistik Pendidikan Tinggi 2017*.

- Kemenristekdikti. 2018. Statistik Pendidikan Tinggi (Higher Educational Statistical Year Book). Pusdatin Kemenristekdikti. <https://pddikti.kemdikbud.go.id/asset/data/publikasi/Statistik Pendidikan Tinggi Indonesia 2018.pdf>
- Kemenristekdikti. 2019. Statistik Pendidikan Tinggi (Higher Education Statistics) 2019. Pusdatin Kemenristekdikti. <https://pddikti.kemdikbud.go.id/asset/data/publikasi/Statistik Pendidikan Tinggi Indonesia 2019.pdf>
- Khasanah, A. U., dan Harwati. 2017. A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques. IOP Conference Series: Materials Science and Engineering, 215(1). 1-7 <https://doi.org/10.1088/1757-899X/215/1/012036>
- Kuh, G. D., Kinzie, J., Buckley, J. A., Bridges, B. K., dan Hayek, J. C. 2006. What Matters to Student Success : A Review of the Literature. In Commissioned Report for the National Symposium on Postsecondary Student Success Spearheading a Dialog on Student Success. 18(7). 1-156
- Kumar, S. A., dan Vijayalakshmi, M. N. 2012. Appraising the Significance of Self Regulated Learning in Higher Education Using Neural Networks. International Journal of Engineering Research and Development, 1(1), 9–15.
- Kupka, J., dan Tomanová, I. 2012. Dependencies among attributes given by fuzzy confirmation measures. Expert Systems with Applications. 39, 7591–7599. <https://doi.org/10.1016/j.eswa.2011.11.125>
- Larose, D. T. 2005. Discovering Knowledge In Data : An Introduction to Data Mining. John Wiley dan Sons, Inc.
- Li, D., Lin, L., dan Peng, L. 2014. Improving learning accuracy by using synthetic samples for small datasets with non-linear attribute dependency. Decision Support Systems, 59, 286–295. <https://doi.org/10.1016/j.dss.2013.12.007>
- Liu, H., dan Yu, L. 2005. Toward Integrating Feature Selection Algorithms for Classification and Clustering. IEEE Transactions On Knowledge And Data Engineering, 17(4), 491–500. <https://doi.org/10.1117/12.942023>
- Manikandan, G., dan Abirami, S. 2021. An efficient feature selection framework based on information theory for high dimensional data. Applied Soft Computing, 7.1-25
- Martinez-torres, M. R., dan Toral, S. L. 2019. A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. Tourism Management. 5. 393–403. <https://doi.org/10.1016/j.tourman.2019.06.003>
- Mathkour, H. 2009. Formalization and Verification of Relational Database Normal Forms Using the Gamma Framework. Journal of King Saud University - Computer and Information Sciences. 21. 1–11. [https://doi.org/10.1016/S1319-1578\(09\)80001-3](https://doi.org/10.1016/S1319-1578(09)80001-3)
- Mayilvaganan, M., dan Kalpanadevi, D. 2015. Comparison of classification techniques for predicting the cognitive skill of students in education environment. IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 4, 113–118. <https://doi.org/10.1109/ICCIC.2014.7238346>
- Mesarić, J., dan Šebalj, D. 2016. Decision trees for predicting the academic success of students. Croatian Operational Research Review. 7, 367–388.

- <https://doi.org/10.17535/crorr.2016.0025>
- Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., dan Punch, W. F. 2003. Predicting student performance: An application of data mining methods with an educational web-based system. *Proceedings - Frontiers in Education Conference*, 1, 1–6.
- Mishra, T., Kumar, D., dan Gupta, S. 2014. Mining students' data for prediction performance. *International Conference on Advanced Computing and Communication Technologies (ACCT)*. 8. 255–262. <https://doi.org/10.1109/ACCT.2014.105>
- Mohamed, M. H., dan Waguih, H. M. 2017. Early Prediction of Student Success Using a Data Mining Classification Technique. *International Journal of Science and Research*. 6(10). 126–131. <https://doi.org/10.21275/ART20177029>
- Mueen, A. 2016. Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. *International Journal of Modern Education and Computer Science*. 11. 36-42. <https://doi.org/10.5815/ijmecs.2016.11.05>
- Musau, O. M., Omieno, K., dan Angulu, R. 2019. Towards Prediction of Students' Academic Performance in Secondary School Using Decision Trees. *International Journal of Research and Innovation in Applied Science (IJRIAS)*, 4(10), 85–89.
- Namoun, A., dan Alshantiti, A. 2021. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*. 11(1), 1–28. <https://doi.org/10.3390/app11010237>
- Natek, S., dan Zwilling, M. 2014. Student data mining solution-knowledge management system related to higher education institutions. *Expert Systems with Applications*. 41(14). 6400–6407. <https://doi.org/10.1016/j.eswa.2014.04.024>
- National Commission for Academic Accreditation dan Assessment. 2015. *The System For Quality Assurance And Accreditation. Handbook for Quality Assurance And Accreditation*. 10. 3-288
- Nuankaew, W., dan Thongkam, J. 2020. Improving Student Academic Performance Prediction Models using Feature Selection. *17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. 9. 392–395. <https://doi.org/10.1109/ECTI-CON49241.2020.9158286>
- Oladokun, V. O., Adebajo, A. T., dan Charles-Owaba, O. E. 2008. Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course. *The Pasific Journal of Science and Technology*, 9(1), 71–79.
- Olmedilla, M., Send, H., dan Toral, S. L. 2019. Technological Forecasting dan Social Change Identification of the unique attributes and topics within Smart Things Open Innovation Communities. *Technological Forecasting dan Social Change*, 146(5), 133–147. <https://doi.org/10.1016/j.techfore.2019.05.004>
- Osmanbegovic, E., dan Suljic, M. 2012. Data Mining Approach for Predicting Student Performance. *Journal of Economics and Business*. 10. 3-12
- Priyasadie, N., dan Isa, S. M. 2021. Educational Data Mining in Predicting Student Final Grades on Standardized Indonesia Data Pokok Pendidikan Data Set. *International Journal of Advanced Computer Science and Applications*,

- 12(12), 212–216. <https://doi.org/10.14569/IJACSA.2021.0121227>
- Punlumjeak, W., dan Rachburee, N. 2015. A comparative study of feature selection techniques for classify student performance. 7th International Conference on Information Technology and Electrical Engineering: Envisioning the Trend of Computer, Information and Engineering (ICITEE). 7. 425–429. <https://doi.org/10.1109/ICITEED.2015.7408984>
- Putpuek, N., Rojanaprasert, N., Atcharyachanvanich, K., dan Thamrongthanyawong, T. 2018. Comparative Study of Prediction Models for Final GPA Score : A Case Study of Rajabhat Rajanagarindra University. IEEE Computer Society. 6. 92–97.
- Quadri, M., dan Kalyankar, D. 2010. Drop out feature of student data for academic performance using decision tree techniques. Global Journal of Computer. 10(2).2–5. <http://computerresearch.org/stpr/index.php/gjst/article/viewArticle/128>
- Rachburee, N., dan Punlumjeak, W. 2015. A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining. International Conference on Information Technology and Electrical Engineering: Envisioning the Trend of Computer, Information and Engineering (ICITEE). 4. 420–424. <https://doi.org/10.1109/ICITEED.2015.7408983>
- Rahman, L., Setiawan, N. A., dan Permanasari, A. E. 2018. Feature selection methods in improving accuracy of classifying students' academic performance. International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 1(1), 267–271. <https://doi.org/10.1109/ICITISEE.2017.8285509>
- Ramesh, V., Parkavi, P., dan Ramar, K. 2013. Predicting Student Performance: A Statistical and Data Mining Approach. International Journal of Computer Applications, 63(8), 35–39. <https://doi.org/10.5120/10489-5242>
- Remeseiro, B., dan Bolon-Canedo, V. 2019. A review of feature selection methods in medical applications. Computers in Biology and Medicine, 112(7), 1-9. <https://doi.org/10.1016/j.combiomed.2019.103375>
- Roblot, T., dan Link, S. 2018. Data & Knowledge Engineering Cardinality constraints and functional dependencies over possibilistic data. Data & Knowledge Engineering, 117(4), 339–358. <https://doi.org/10.1016/j.datak.2018.04.005>
- Romero, C., Ventura, S., Espejo, P. G., dan Hervás, C. 2008. Data mining algorithms to classify students. Educational Data Mining 1st International Conference on Educational Data Mining. 3. 8–17.
- Ron Kohavi, dan John, G. H. 1997. Wrappers for feature subset selection. Artificial Intelligence 97. 11. 654–678. https://doi.org/10.1007/978-3-642-39038-8_27
- Sembiring, S., Zarlis, M., Hartama, D., Ramliana, S., dan Wani, E. 2011. Prediction of Student Academic Performance By an Application of Data Mining Techniques. Management and Artificial Intelligence. 6(1). 110–114.
- Shahiri, A. M., Husain, W., dan Rashid, N. A. 2015. A Review on Predicting Student's Performance Using Data Mining Techniques. Procedia Computer Science, 72, 414–422. <https://doi.org/10.1016/j.procs.2015.12.157>
- Shahiri, A. M., Husain, W., dan Rashid, N. A. 2017. Handling high dimensional educational data using feature selection techniques. Journal of

- Telecommunication, Electronic and Computer Engineering. 9(2). 89–93.
- Shiva Darshan, S. L., dan Jaidhar, C. D. 2018. Performance Evaluation of Filter-based Feature Selection Techniques in Classifying Portable Executable Files. *Procedia Computer Science*. 125. 346–356. <https://doi.org/10.1016/j.procs.2017.12.046>
- Shroff, K. P., dan Maheta, H. H. 2015. A comparative study of various feature selection techniques in high-dimensional data set to improve classification accuracy. *International Conference on Computer Communication and Informatics (ICCCI)*. 1. 6-12. <https://doi.org/10.1109/ICCCI.2015.7218098>
- Silberschatz, A., Korth, H. F., dan Sudarshan, S. 2010. *Database System Concepts* 6th edition. McGraw-Hill. <https://doi.org/10.1080/09638280500030605>
- Singh, W. 2016. Comparative Analysis of Classification Techniques for Predicting Computer Engineering Students ' Academic Performance. *International Journal of Advanced Research in Computer Science*. 7(6). 31-36
- Sivasakthi, M. 2018. Classification and prediction based data mining algorithms to predict students' introductory programming performance. *International Conference on Inventive Computing and Informatics (ICICI)*. 9. 346–350. <https://doi.org/10.1109/ICICI.2017.8365371>
- Sokkhey, P., dan Okazaki, T. 2020. Study on dominant factor for academic performance prediction using feature selection methods. *International Journal of Advanced Computer Science and Applications*, 11(8), 492–502. <https://doi.org/10.14569/IJACSA.2020.0110862>
- Srivastava, M. S., Joshi, M. N., dan Gaur, M. M. 2013. A Review Paper on Feature Selection Methodologies and Their Applications. *International Journal of Computer Science and Network Security*. 14(5). 78–81.
- Tair, M. M. A., dan El-Halees, A. M. 2012. Mining Educational Data to Analyze Students' Performance: A Case Study of Mawuli School, Ho. *International Journal of Information and Communication Technology Research*. 2(2). 140–146. <https://doi.org/10.38124/ijisrt20may635>
- Tour, A., dan Mathkour, H. 2008. An application of Gamma Formalism in Database Design. *International Conference on Computer and Communication Engineering*. 5. 974–977.
- Triayudi, A., dan Fitri, I. 2021. Comparison Of The Feature Selection Algorithm In Educational Data Mining. *Telkomnika (Telecommunication Computing Electronics and Control)*. 19(6). 1865–1871. <https://doi.org/10.12928/TELKOMNIKA.v19i6.21594>
- Usamah, Buniyamin, N., Arsad, P. M., dan Kassim, R. A. 2013. An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention. *International Conference on Engineering Education: Aligning Engineering Education with Industrial Needs for Nation Development (ICEED)*. 8. 126–130. <https://doi.org/10.1109/ICEED.2013.6908316>
- Vassiliou, Y. 2015. Functional Dependencies and Incomplete Information. *Center for Research on Information Systems*. 4. 1-12
- Venkatesh, B., dan Anuradha, J. 2019. A Review of Feature Selection and Its Methods. *Cybernetics And Information Technologies*. 19(1), 3–26. <https://doi.org/10.2478/cait-2019-0001>
- Wald, R., Khoshgoftaar, T., dan Napolitano, A. 2013. Comparison of stability for

- different families of filter-based and wrapper-based feature selection. *International Conference on Machine Learning and Applications (ICMLA)*. 2. 457–464. <https://doi.org/10.1109/ICMLA.2013.162>
- Wang, T., dan Mitrovic, A. 2002. Using Neural Networks to Predict Student's Performance. *International Conference on Computers in Education*. 3. 15-20
- Wook, M., Yahaya, Y. H., Wahab, N., Isa, M. R. M., Awang, N. F., dan Seong, H. Y. 2009. Predicting NDUM student's academic performance using data mining techniques. *International Conference on Computer and Electrical Engineering (ICCEE)*. 2. 357–361. <https://doi.org/10.1109/ICCEE.2009.168>
- Wu, W., dan Mao, W. 2022. An Efficient and Scalable Algorithm to Mine Functional Dependencies from Distributed Big Data. *Sensors*. 22(10), 1–19. <https://doi.org/10.3390/s22103856>
- Xue, B., Zhang, M., Browne, W. N., dan Yao, X. 2016. A Survey on Evolutionary Computation Approaches to Feature Selection. *IEEE Transactions on Evolutionary Computation*. 20(4). 606–626. <https://doi.org/10.1109/TEVC.2015.2504420>
- Yamaguchi, D. 2009. *International Journal of Approximate Reasoning*. *International Journal of Approximate Reasoning*. 51(1). 89–98. <https://doi.org/10.1016/j.ijar.2009.08.002>
- Yassein, N. A., M Helali, R. G., dan Mohomad, S. B. 2017. Predicting Student Academic Performance in KSA using Data Mining Techniques. *Journal of Information Technology dan Software Engineering*. 07(05). <https://doi.org/10.4172/2165-7866.1000213>
- York, T. T., Gibson, C., dan Rankin, S. 2015. Defining and measuring academic success. *Practical Assessment, Research and Evaluation*. 20(5). 1–20.
- Zaffar, M., Hashmani, M. A., dan Savita, K. S. 2018. Performance analysis of feature selection algorithm for educational data mining. *IEEE Conference on Big Data and Analytics (ICBDA)*. 1. 7–12. <https://doi.org/10.1109/ICBDAA.2017.8284099>
- Zaffar, M., Hashmani, M. A., Savita, K. S., dan Rizvi, S. S. H. 2018. A study of feature selection algorithms for predicting students academic performance. *International Journal of Advanced Computer Science and Applications*. 9(5). 541–549. <https://doi.org/10.14569/IJACSA.2018.090569>
- Zheng, W. 2018. A Comparative Study of Feature Selection Methods. *International Journal on Natural Language Computing*. 7(5). 1–9. <https://doi.org/10.5121/ijnlc.2018.7501>
- Zhou, H. F., Zhang, Y., Zhang, Y. J., dan Liu, H. J. 2019. Feature selection based on conditional mutual information: minimum conditional relevance and minimum conditional redundancy. *Applied Intelligence*, 49(3), 883–896. <https://doi.org/10.1007/s10489-018-1305-0>