

**IMPLEMENTASI BERBAGAI METODE  
KECERDASAN BUATAN (*ARTIFICIAL INTELLIGENCE*)  
PADA MASALAH GANGGUAN KEPERIBADIAN  
(*NARCISSISTIC PERSONALITY DISORDER: NPD*)**

**Disertasi**

**Oleh**

**HENI SULISTIANI  
NPM. 1737061011**



**PROGRAM STUDI DOKTOR (S3) MIPA  
PASCASARJANA  
UNIVERSITAS LAMPUNG  
2024**

## ABSTRAK

# IMPLEMENTASI BERBAGAI METODE KECERDASAN BUATAN (*ARTIFICIAL INTELLIGENCE*) PADA MASALAH GANGGUAN KEPRIBADIAN (*NARCISSISTIC PERSONALITY DISORDER: NPD*)

Oleh

HENI SULISTIANI

Gangguan mental (*mental disorder*) merupakan kondisi kesehatan yang memengaruhi pola pikir, perasaan, perilaku, suasana hati, atau campuran dari berbagai kondisi seseorang. Salah satu jenis gangguan mental yang saat ini sedang ramai dibahas yaitu gangguan kepribadian narsistik. Penderita gangguan kepribadian narsistik yang tidak segera ditangani, akan berdampak negatif pada kehidupan individu dan hubungan mereka dengan orang lain, bahkan beresiko pada perilaku bunuh diri. Proses yang dilakukan untuk mendiagnosa pasien gangguan kepribadian narsistik saat ini hanya menggunakan metode tradisional, hanya berdasarkan laporan verbal pasien dan sesi wawancara yang panjang dengan profesional medis, yang hasilnya bisa subjektif, mahal, dan membutuhkan waktu yang lama. Penelitian yang telah dikembangkan saat ini juga hanya berfokus pada pengembangan aplikasi untuk menyelesaikan permasalahan atau membantu dalam mendiagnosa jenis gangguan kepribadian. Belum ditemukan penelitian yang mengembangkan model teknik pemilihan fitur dan mengevaluasinya untuk memperoleh fitur-fitur yang relevan dengan kelas sehingga mampu meningkatkan akurasi metode *machine learning*.

Tujuan dilakukannya penelitian ini antara lain melakukan studi literatur dan analisis bibliometrik tren kecerdasan buatan (*artificial intelligence: AI*) untuk memprediksi gangguan kepribadian dengan menggunakan teknik *systematic literature review* (SLR) dengan visualisasi menggunakan VOSviewer; mengevaluasi teknik pemilihan fitur (*feature selection*) guna menghasilkan fitur-fitur yang relevan dan dapat meningkatkan kinerja metode *machine learning*; mengevaluasi dan melakukan perbandingan kinerja berbagai metode *machine learning* dalam memprediksi *narcissistic personality disorder* dan membangun model *synthetic minority oversampling technique* (SMOTE) untuk mengatasi ketidakseimbangan kelas. Proses pemilihan fitur menggunakan Teknik *information gain* dan *gain ratio*. Sedangkan pada metode *machine learning* menerapkan tiga metode klasifikasi yaitu *random forest classifier*, *support vector machine* dan *naive bayes*.

Hasil penelitian pertama berdasarkan *literature review* menunjukkan bahwa terdapat 22 artikel terindeks Scopus dari 1.022 artikel di Google Scholar dan dianalisis menggunakan teknik bibliografi dan visualisasi menggunakan VOSviewer. Dari proses tersebut, ditemukan 138 kata kunci yang dibagi kedalam tujuh kluster. Hasil penelitian kedua menunjukkan bahwa dari proses evaluasi kinerja, metode *random forest classifier*

memiliki nilai akurasi tertinggi bila dibandingkan dengan metode *support vector machine* dan *naive bayes* dalam ketiga skenario (menggunakan semua fitur, seleksi fitur *information gain*, dan seleksi fitur *gain ratio*). Nilai akurasi tersebut secara berturut-turut sebesar 99,93%, 99,96%, dan 100%. Selain itu, diperoleh hasil bahwa metode Naive Bayes dengan *information gain* memiliki waktu pemrosesan tercepat; hanya membutuhkan waktu 0,22 detik. Hasil penelitian ketiga dalam penerapan teknik SMOTE untuk mengatasi ketidakseimbangan menunjukkan bahwa teknik tersebut mampu meningkatkan nilai akurasi pengklasifikasi (*Naive Bayes*). Nilai akurasi sebelum menerapkan teknik SMOTE hanya sebesar 86,52%. Setelah menerapkan teknik SMOTE, akurasi meningkat menjadi 91,96%. Jadi hasil penelitian ini menunjukkan bahwa metode *machine learning* yang dikembangkan mampu memberikan nilai akurasi yang mencapai lebih dari 90%. Hal ini berarti metode yang diterapkan dapat dengan baik melakukan prediksi data *narcissistic personality disorder*.

Kata kunci: Gangguan Kepribadian, Gangguan Mental, Machine Learning, Narcissistic Personality Disorder, Seleksi Fitur

## **ABSTRACT**

### **IMPLEMENTATION OF VARIOUS METHODS ARTIFICIAL INTELLIGENCE ON THE PROBLEM OF PERSONALITY DISORDERS (NARCISSISTIC PERSONALITY DISORDER: NPD)**

**By**

**HENI SULISTIANI**

Mental disorders are health conditions that affect a person's thinking patterns, feelings, behavior, mood, or a mixture of various conditions. One type of mental disorder that is currently being widely discussed is narcissistic personality disorder. Patients with narcissistic personality disorder who are not treated immediately will have a negative impact on the individual's life and their relationships with other people, even at risk of suicidal behavior. The process used to diagnose narcissistic personality disorder patients currently uses traditional methods, based only on the patient's verbal report and long interview sessions with medical professionals, the results of which can be subjective, expensive, and take a long time. The research that has been developed currently only focuses on developing applications to solve problems or help in diagnosing types of personality disorders. No research has been found that develops feature selection technique models and evaluates them to obtain class-relevant features so as to increase the accuracy of machine learning methods.

The aims of this research include conducting a literature study and bibliometric analysis of artificial intelligence (AI) trends to predict personality disorders using systematic literature review (SLR) techniques with visualization using VOSviewer; evaluate feature selection techniques to produce relevant features and can improve the performance of machine learning methods; evaluating and comparing the performance of various machine learning methods in predicting narcissistic personality disorder and building a synthetic minority oversampling technique (SMOTE) model to overcome class imbalance. The feature selection process uses information gain and gain ratio techniques. Meanwhile, the machine learning method applies three classification methods, namely random forest classifier, support vector machine and naive Bayes.

The results of the first research based on a literature review showed that there were 22 articles indexed by Scopus out of 1,022 articles on Google Scholar and analyzed using bibliographic and visualization techniques using VOSviewer. From this process, 138 keywords were found which were divided into seven clusters. The results of the second research show that from the performance evaluation process, the random forest classifier method has the highest accuracy value when compared with the support vector machine and naive Bayes methods in the three scenarios (using all features, information

gain feature selection, and gain ratio feature selection). The accuracy values are 99.93%, 99.96% and 100% respectively. In addition, the results showed that the Naive Bayes method with information gain had the fastest processing time; it only takes 0.22 seconds. The results of the third research in applying the SMOTE technique to overcome imbalances show that this technique is able to increase the accuracy value of the classifier (Naive Bayes). The accuracy value before applying the SMOTE technique was only 86.52%. After applying the SMOTE technique, the accuracy increased to 91.96%. So the results of this research show that the machine learning method developed is able to provide accuracy values that reach more than 90%. This means that the method applied can predict narcissistic personality disorder data well.

Keyword: Personality Disorder, Mental Health, Machine Learning, Narcissistic Personality Disorder, Feature Selection

## MENYETUJUI

Judul Disertasi : Implementasi Berbagai Metode Kecerdasan Buatan (*Artificial Intelligence*) Pada Masalah Gangguan Kepribadian (*Narcissistic Personality Disorder: NPD*)

Nama Mahasiswa : Heni Sulistiani

Nomor Pokok Mahasiswa : 1737061011


Program Studi : Doktor MIPA

Fakultas : Matematika dan Ilmu Pengetahuan Alam

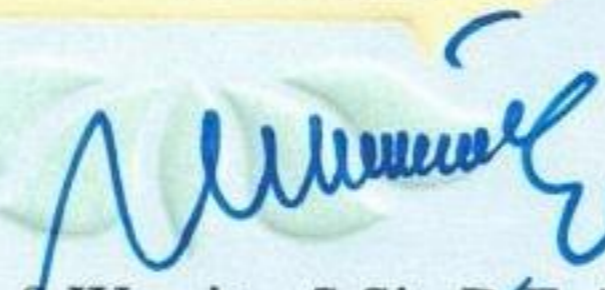
  
Promotor,  
Prof. Admi Syarif, Ph.D.  
NIP. 196701031992031003

### Komisi Promotor


Ko-Promotor 1

  
Dr. dr. Khairun Nisa Berawi, M.Kes., AIFO-K., FISCM.  
NIP. 196406161989021001

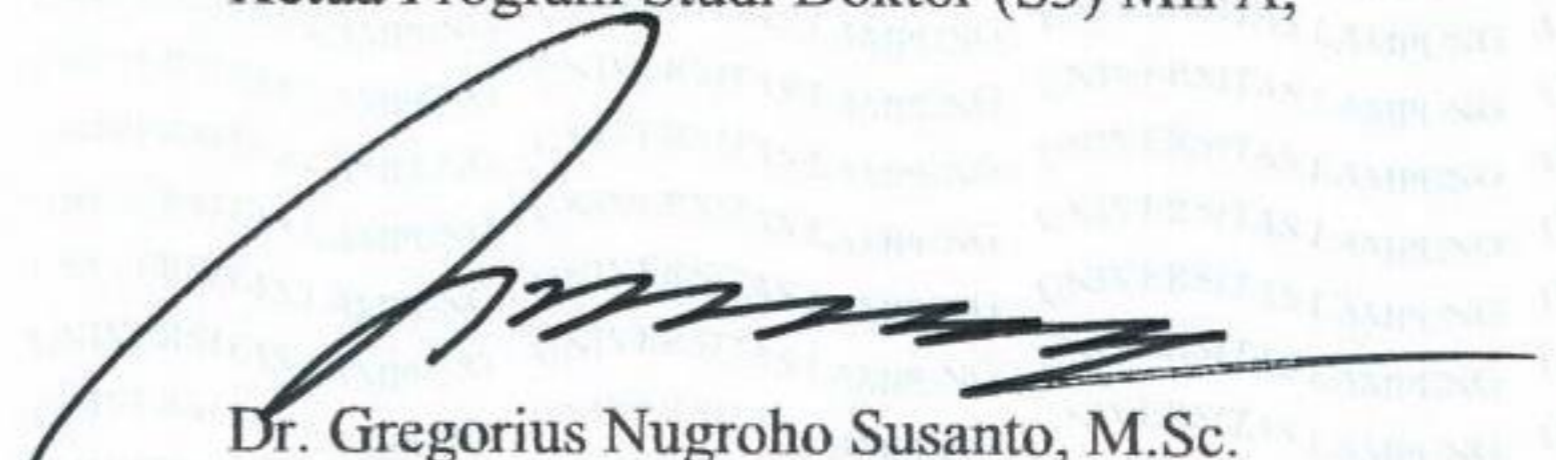
Ko-Promotor 2

  
Prof. Warsito, S.Si., D.E.A., Ph.D.  
NIP. 197102121995121001

Direktur Pascasarjana,

  
Prof. Dr. Ir. Murhadi, M.Si.  
NIP. 196403261989021001

Ketua Program Studi Doktor (S3) MIPA,

  
Dr. Gregorius Nugroho Susanto, M.Sc.  
NIP. 196103111988031001

## MENGESAHKAN

### 1. Tim Penguji

Promotor : Prof. Admi Syarif. Ph.D. (.....)

Ko-Promotor 1 : Dr. dr. Khairun Nisa Berawi, S.Ked., (.....)  
M. Kes., AIFO-K., FISC.M.

Ko-Promotor 2 : Prof. Warsito, S.Si., D.E.A., Ph.D. (.....)

Penguji Internal 1 : Dr. rer.nat. Akmal Junaidi, M.Sc. (.....)

Penguji Internal 2 : Prof. Dr. Hendri Busman, M.Biomed. (.....)

Penguji Eksternal : RZ. Abdul Aziz, ST., MT., Ph.D. (.....)

### 2. Dekan Fakultas Matematika IPA



Dr. Eng. Heri Satria, S.Si., M.Si.

NIP. 197110012005011002

Tanggal lulus ujian : 22 Januari 2024

## LEMBAR IDENTITAS DAN PENGESAHAN

- A. Judul Disertasi : Implementasi Berbagai Metode Kecerdasan Buatan (*Artificial Intelligence*) pada Masalah Gangguan Kepribadian (*Narcissistic Personality Disorder: NPD*)
- B. Bidang Ilmu : Ilmu Komputer
1. Ketua
- a. Nama : Heni Sulistiani
- b. Gol/Pangkat/NPM : III.C/ Penata Tk.I /1737061011
- c. Jabatan Fungsional : Lektor
- d. Fakultas/Bagian : Matematika dan Ilmu Pengetahuan Alam / Ilmu Komputer
- e. Universitas : Lampung
2. Anggota Peneliti : 3 (tiga) orang
3. Lokasi Penelitian : Kota Bandar Lampung

Bandar Lampung, 22 Januari 2024

Mengetahui,  
Dekan Fakultas MIPA Unila

Peneliti



Dr. Eng. Heri Satria, S.Si., M.Si.  
NIP. 197110012005011002

Heni Sulistiani  
NPM. 1737061011



## RIWAYAT HIDUP



**Heni Sulistiani** merupakan sosok nama penulis dari Disertasi ini. Penulis lahir dari pasangan Bapak Haryanto dan Ibu Suryani sebagai anak pertama dari dua bersaudara. Penulis dilahirkan di Polaman, Kabupaten Pringsewu pada tanggal 12 Oktober 1986. Tahun 1998 penulis menyelesaikan pendidikan tingkat dasar di SD N 2 Pringsewu. Tahun 2001

penulis menyelesaikan pendidikan tingkat menengah di SLTP N 1 Pringsewu dan tahun 2004 penulis menyelesaikan pendidikan tingkat atas di SMA N 1 Pringsewu.

Tahun 2004 penulis juga mendaftarkan diri di AMIK Teknokrat untuk melanjutkan studi ke jenjang Diploma 3 program studi Komputerisasi Akuntansi. Selama menjadi mahasiswa di Diploma 3, penulis pernah menjadi asisten dosen laboratorium komputer. Tahun 2006 penulis melakukan Praktik Kerja Lapangan (PKL) di Tunas Mandiri Finance Cabang Pringsewu. Tahun 2007 penulis berhasil menyelesaikan studi Diploma 3 dan melanjutkan studi program sarjana (S-1) di STMIK Teknokrat program studi Teknik Informatika. Tahun 2009 penulis melakukan Praktik Kerja Lapangan (PKL) untuk syarat kelulusan program sarjana di PT Semen Baturaja Persero Pabrik Panjang. Selain itu, selama menjadi mahasiswa, penulis juga pernah mengikuti seminar nasional sebagai pemakalah di Seminar Nasional Aplikasi Teknologi Informasi (SNATI) dengan judul “Rekayasa Penawaran Produk Asuransi Secara Online pada PT. AIG Life Lampung”.

Penulis lulus dari program sarjana pada tahun 2012 dengan judul skripsi “Analisis Strategi Promosi menggunakan Algoritma C45 pada PT Semen Baturaja (Persero) Pabrik Panjang”. Setelah lulus program sarjana di tahun 2012, penulis bekerja di Perguruan Teknokrat sebagai staff kemahasiswaan. Tahun 2014 penulis melanjutkan studi ke jenjang Magister (S-2) di Institut Teknologi Sepuluh Nopember program studi Sistem Informasi. Penulis lulus dari program magister tahun 2016 dengan judul tesis “Pemilihan Fitur untuk Klasifikasi Loyalitas Pelanggan terhadap Merk Produk Fast Moving Consumer”.

## **HALAMAN PERSEMBAHAN**

Alhamdulillah Rabbil Aalamin, sujud serta syukur kepada Allah SWT. Terimakasih atas karunia-Mu yang telah memberikan kemudahan dan kelancaran sehingga disertai ini dapat terselesaikan dengan baik.

Kupersembahkan karya sederhana ini kepada orang yang sangat kukasihi dan kusayangi.

### **Ibu dan Bapak Tercinta**

Sebagai tanda bakti, hormat dan rasa terima kasih yang tiada terhingga kupersembahkan karya kecil ini kepada Ibu (Suryani) dan Bapak (Haryanto) yang telah memberikan kasih sayang, secara dukungan, ridho, dan cinta kasih yang tiada terhingga yang tiada mungkin dapat kubalas hanya dengan selembar kertas yang bertuliskan kata persembahan. Semoga ini menjadi langkah awal untuk membuat Ibu dan Bapak bahagia karena kusadar, selama ini belum bisa berbuat lebih. Untuk Ibu dan Bapak yang selalu membuatku termotivasi dan selalu menyirami kasih sayang, selalu mendoakanku, selalu menasehatiku serta selalu meridhoiku melakukan hal yang lebih baik,

### **Suamiku dan Anakku Tersayang**

Terima kasih kuucapkan kepada suamiku tersayang yang telah sabar, mencurahkan seluruh kasih sayang, selalu memberikan dukungan, do'a, cinta kasih yang tak terhingga, selalu memberikan motivasi dan menjadi pendengar yang baik atas segala keluh kesah.

### **Adik-Adiku dan Ibu Mertuaku Terkasih**

Terima kasih telah memberikan semangat dan inspirasi dalam menyelesaikan Tugas Akhir ini. Semoga doa dan semua hal yang terbaik yang engkau berikan menjadikan ku orang yang baik pula.

## KATA PENGANTAR

Puji dan syukur penulis ucapkan kehadiran Allah SWT, atas segala berkah, rahmat, dan karunia-Nya yang telah memberikan ilmu pengetahuan, pengalaman, kekuatan, kesabaran, dan kesempatan kepada penulis sehingga mampu menyelesaikan disertasi ini. Akan tetapi sesungguhnya penulis menyadari bahwa tanpa bantuan dan dukungan dari berbagai pihak, maka penyusunan disertasi ini tidak dapat berjalan dengan baik. Hingga selesainya penulisan disertasi ini telah banyak menerima bantuan waktu, tenaga dan pikiran dari banyak pihak. Sehubungan dengan itu, maka pada kesempatan ini perkenankanlah penulis menyampaikan terima kasih sebesar-besarnya kepada:

1. Ibu Prof. Dr. Ir. Lusmeilia Afriani, D.E.A., IPM., ASEAN Eng. Selaku Rektor Universitas Lampung.
2. Bapak Dr. Eng. Heri Satria, S.Si., M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.
3. Bapak Dr. Gregorius Nugroho Susanto, M.Sc. selaku ketua Program Studi Doktor MIPA, Universitas Lampung.
4. Bapak Prof. Admi Syarif, Ph.D. selaku promotor yang selalu membimbing dengan baik, memberikan masukan dan pemecahan masalah yang tepat, dan membimbing dengan sepenuh hati. Hingga disertasi ini dapat terselesaikan dengan baik.
5. Ibu Dr. dr. Khairun Nisa Berawi, M.Kes., AIFO-K., FISC.M. Selaku ko-promotor 1 yang selalu memberikan motivasi, mengingatkan saya untuk selalu teliti dan tepat waktu, dan memberikan arahan yang baik, sehingga penyusunan disertasi ini dapat berjalan dengan lancar.
6. Bapak Prof. Warsito, S.Si., D.E.A., Ph.D. selaku ko-promotor 2 yang telah meluangkan waktunya untuk memberikan masukan dan sabar membimbing hingga terselesaikannya disertasi ini.

7. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc. selaku pembahas 1 yang telah memberikan saran dan masukan untuk proses perbaikan disertasi ini.
8. Bapak Prof. Dr. Hendri Busman, M.Biomed. selaku pembahas 2 yang telah memberikan saran dan masukan untuk proses perbaikan disertasi ini.
9. Ibu Hj. Hernaini, SS., M.Pd. selaku Pembina Yayasan Pendidikan Teknokrat yang telah memberikan dukungan dan motivasi hingga terselesaikannya disertasi ini.
10. Ibu Dewi Sukmasari, SE., MSA., CA., Akt. Selaku Ketua Yayasan Pendidikan Teknokrat yang telah memberikan dukungan dan motivasi hingga terselesaikannya disertasi ini.
11. Bapak Dr. HM. Nasrullah Yusuf, SE., MBA. selaku Rektor Universitas Teknokrat Indonesia yang telah memberikan dukungan dan motivasi kepada penulis dalam penyelesaian disertasi.
12. Bapak Dr. H. Mahathir Muhammad, SE., MM. selaku Dekan Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia yang selalu memberikan dukungan, arahan, motivasi dan selalu mengingatkan agar penulis dapat menyelesaikan disertasi dengan baik.
13. Teman-teman seperjuangan di Prodi Doktor MIPA Unila khususnya Angkatan 2017, jangan menyerah dan lelah untuk menyelesaikan disertasi, terus semangat untuk menggapai impian lulus program doktoral.
14. Sahabat-Sahabat terkasih dan rekan kerja di Universitas Teknokrat Indonesia yang telah memberikan dukungan dan semangat dalam penyelesaian disertasi ini.

Semoga Allah SWT, memberikan balasan dengan segala kebaikan dunia dan ahirah atas keikhlasan dan dan kebaikan semua pihak yang telah diberikan kepada penulis. Harapan penulis semoga disertasi ini dapat bermanfaat bagi semua pihak yang membacanya, khususnya pengembangan untuk ilmu komputer.

Penulis menyadari bahwa masih banyak kekurangan dan ketidak sempurnaan didalam penelitian disertasi ini. Untuk itu penulis mengharapkan kritik dan saran untuk menyempurnakan dimasa yang akan datang. Penulis juga mengucapkan terima kasih kepada seluruh pihak atas perhatian dan pemberian semangat selama proses penyelesaian disertasi.

Bandar Lampung, November 2023

**Heni Sulistiani**

## DAFTAR ISI

	Halaman
ABSTRAK .....	i
ABSTRACT .....	iii
LEMBAR PERSETUJUN .....	v
RIWAYAT HIDUP .....	v
HALAMAN PERSEMBAHAN .....	vii
KATA PENGANTAR .....	ix
DAFTAR ISI .....	xii
DAFTAR GAMBAR .....	xiv
DAFTAR TABEL .....	xv
DAFTAR LAMPIRAN .....	xvi
I. PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	5
1.3 Tujuan Penelitian .....	5
1.4 Manfaat Penelitian .....	6
1.5 Batasan Masalah .....	6
1.6 Kontribusi Penelitian .....	7
1.7 Keterbaruan ( <i>Novelty</i> ) .....	7
II. TINJAUAN PUSTAKA .....	9
2.1. Studi Literatur .....	9
2.2. Gangguan Mental ( <i>Mental Disorder</i> ) .....	11
2.3. Kecerdasan Buatan ( <i>Artificial Intelligence</i> ) .....	16
2.4. Mesin Pembelajaran ( <i>Machine Learning</i> ) .....	29
2.5. Kecerdasan Buatan untuk Prediksi <i>Narcisistic Personality Disorder</i> .....	40
2.6. Database Google Scholar .....	42
2.7. Aplikasi Publish or Perish .....	43
2.8. Aplikasi VOSviewer .....	44
III. METODOLOGI PENELITIAN .....	43
3.1. Tahapan Penelitian .....	43
3.2. Diagram Alir Penelitian .....	48
3.2.1. Diagram Alir Penelitian Studi Literature Review .....	49
3.2.2. Diagram Alir Penelitian Proses Evaluasi Kinerja Teknik Pemilihan Fitur pada Berbagai Metode <i>Machine Learning</i> ...	54
3.2.3. Diagram Alir Penelitian Implementasi SMOTE-Naive Bayes untuk Mengatasi Ketidakseimbangan Kelas .....	71
IV. HASIL DAN PEMBAHASAN .....	74
4.1. Analisis Bibliometrik menggunakan Teknik <i>Systematic Literature     Review (SLR) Kecerdasan Buatan pada Gangguan Mental</i> .....	74
4.1.1. Pendekatan Kecerdasan Buatan .....	76



4.1.2. Perkembangan Publikasi Penelitian Penerapan <i>Machine Learning/ Artificial Intelligence</i> untuk Prediksi Gangguan Kepribadian .....	86
4.1.3. Top 5 Publisher .....	87
4.1.4. Analisis Sitasi .....	89
4.1.5. Analisis Kata Kunci ( <i>Co-occurrence</i> ) .....	106
4.2. Evaluasi Kinerja Teknik Pemilihan Fitur pada Berbagai Metode <i>Machine Learning</i> untuk Diagnosa <i>Narcisistic Personality Disorder (NPD)</i> .....	110
4.2.1. Eksperimen Seleksi Fitur .....	114
4.2.2. Eksperimen Klasifikasi .....	118
4.3. Implementasi Teknik SMOTE-Naive Bayes untuk Mengatasi Ketidakseimbangan Kelas pada Prediksi Gangguan <i>Narcisistic Personality Disorder</i> .....	122
4.3.1. Implementasi SMOTE-Naive Bayes .....	126
4.3.2. Perbandingan Kinerja SMOTE-Naive Bayes dengan Naive Bayes .....	128
V. KESIMPULAN .....	116
5.1. Kesimpulan .....	116
5.2. Saran .....	117
VI. DAFTAR PUSTAKA .....	118

## DAFTAR GAMBAR

Gambar	Halaman
1. Tahapan Penelitian .....	44
2. Tampilan dataset <i>narcissistic personality disorder</i> .....	45
3. Diagram alir penelitian.....	48
4. Diagram Alir Proses SLR.....	50
5. Metode Penelusuran menggunakan PoP .....	52
6. Diagram Alir Proses Evaluasi Kinerja Teknik Pemilihan Fitur .....	55
7. Dataset dengan format .csv .....	56
8. Daftar pernyataan pada kuesioner .....	57
9. Prosedur <i>k-fold Cross-Validation</i> .....	70
10. Diagram Alir Penelitian Proses Implementasi SMOTE-Naive Bayes .....	72
11. Tahapan pendekatan kecerdasan buatan dalam prediksi gangguan kepribadian.....	77
12. Komponen sistem pakar .....	80
13. TOP 5 <i>publisher</i> dengan artikel ilmiah terbanyak .....	88
14. <i>Co-occurrence</i> kata kunci pada penelitian kecerdasan buatan untuk gangguan kepribadian.....	107
15. Visualisasi <i>density</i> kata kunci pada bidang kecerdasan buatan .....	108
16. Visualisasi kata kunci teratas untuk bidang kecerdasan buatan.....	109
17. Proses seleksi fitur menggunakan Information Gain .....	115
18. Proses seleksi fitur menggunakan Gain Ratio.....	115
19. Pengaturan nilai <i>threshold</i> pada WEKA .....	116
20. Perbandingan nilai <i>error rate</i> dari metode ML.....	120
21. Perbandingan waktu pemrosesan .....	121
22. Diagram penyebaran data tidak seimbang .....	126
23. Tampilan penerapan Naive Bayes pada Weka.....	128

## DAFTAR TABEL

Tabel	Halaman
1. Penelitian terdahulu.....	9
2. Deskripsi Fitur.....	46
3. <i>Confusion Matrix</i> .....	47
4. Kriteria artikel pada proses <i>literature review</i> .....	51
5. Hasil <i>screening</i> artikel.....	53
6. Deskripsi fitur.....	57
7. <i>Confusion matrix</i> .....	71
8. Perkembangan Publikasi Penelitian tentang Penerapan <i>Machine Learning/ Artificial Intelligence</i> untuk Prediksi Gangguan Kepribadian .....	87
9. Jumlah sitasi pada artikel tentang kecerdasan buatan untuk gangguan kepribadian yang terindeks <i>Scopus Quartile Q1-Q4</i> .....	74
10. Top 10 Kata kunci yang ditemukan pada judul dan abstrak .....	110
11. Ranking hasil seleksi fitur.....	116
12. Hasil perbandingan teknik pemilihan fitur pada metode <i>machine learning</i> .	118
13. Perbandingan data awal dengan data SMOTE.....	127
14. Hasil perbandingan Naive Bayes dengan SMOTE-Naive Bayes .....	129

## DAFTAR LAMPIRAN

Lampiran	Halaman
1. Bukti Artikel Telah terbit pada <i>International Conference on Applied Sciences Mathematics and Informatics (ICASMI 2020)</i> .....	132
2. Sertifikat sebagai presenter di ICASMI 2020 .....	133
3. Bukti Submit pada Jurnal <i>Bulletin of Electrical Engineering and Informatics (BEEI)</i> .....	134
4. Bukti sedang dalam proses editing di Jurnal <i>Bulletin of Electrical Engineering and Informatics (BEEI)</i> .....	134
5. Bukti artikel telah diterima pada Jurnal <i>Bulletin of Electrical Engineering and Informatics (BEEI)</i> .....	135
6. Artikel Penelitian yang Telah Submit di International Journal of Intelligent Systems and Applications in Engineering (IJISAE) .....	136

# I. PENDAHULUAN

Pada bagian ini dipaparkan latar belakang permasalahan dan identifikasi masalah yang menjadi dasar penelitian, rumusan penelitian, tujuan penelitian, manfaat penelitian, kontribusi penelitian dan keterbaruan (*novelty*).

## 1.1 Latar Belakang

Gangguan mental (*mental disorder*) merupakan kondisi kesehatan seseorang yang memengaruhi pola pikir, perasaan, perilaku, suasana hati, atau campuran dari berbagai kondisi seseorang. Kondisi tersebut dapat terjadi sesekali atau dapat berlangsung dalam kurun waktu yang lama (kronis) (Primananda 2022). Peristiwa yang dialami oleh seseorang sangat berpengaruh pada kepribadian dan perilaku seseorang sehingga berdampak pada adanya gangguan mental (Savitrie 2022). Pada tahun 2022, terdapat sekitar satu dari tujuh remaja di dunia yang mengalami gangguan mental (WHO 2022). Gangguan mental yang dialami oleh seseorang sangat beragam antara lain, gangguan kecemasan, gangguan kepribadian, gangguan psikotik, gangguan suasana hati, gangguan makan, gangguan pengendalian impuls dan kecanduan, gangguan obsesif kompulsif (OCD), gangguan stress pasca trauma (PTSD). Gangguan kesehatan mental merupakan masalah yang jarang mendapat perhatian dan penanganan yang serius (Dewi dan Ariana 2021). Hal ini dikarenakan masih banyak masyarakat yang belum mengetahui bahaya gangguan mental, sehingga penderita takut untuk berbicara dan mengungkapkan apa yang sedang dialaminya. Kasus gangguan kepribadian merupakan kasus yang paling sering dijumpai (Sari *et al.* 2016).

Hasil Riset Kesehatan Dasar (Riskesdas) pada tahun 2018, menunjukkan bahwa jumlah penderita gangguan kepribadian (gangguan mental emosional) mencapai 6,1% dari total penduduk di Indonesia ((KemenkesRI) 2019). Seseorang yang

menderita gangguan kepribadian akan memiliki cara berpikir dan perilaku yang tidak sehat dan berbeda dari orang pada umumnya (Teguh *et al.* 2019). Gangguan kepribadian bermula dari munculnya stres yang kemudian berlanjut pada penekanan perasaan-perasaan dan berperilaku tertentu seperti orang yang mengalami stress pada umumnya (Johar *et al.* 2014). Seseorang yang menderita gangguan kepribadian memiliki respons yang benar-benar kaku terhadap situasi pribadi, dalam menjalin hubungan sosial dengan orang lain di lingkungan rumah, sekolah, kantor, atau pekerjaan (Sari *et al.* 2016). Gangguan kepribadian dapat menyebabkan depresi yang berujung pada kasus bunuh diri (Pahlevi *et al.* 2018). Beberapa jenis gangguan kepribadian antara lain gangguan kepribadian ambang (*borderline*), antisosial, narsistik, histrionic, schizoid, paranoid dan lain sebagainya. Salah satu jenis gangguan kepribadian yang sedang ramai dibahas yaitu gangguan kepribadian narsistik.

Pada masyarakat modern, perilaku narsis telah menjadi sebuah endemic seiring dengan perkembangan media sosial yang semakin pesat (Pahlevi *et al.* 2018). Munculnya media sosial seperti *Instagram, Facebook, Twitter, Tik Tok, WhatsApp* dan lain sebagainya dapat memberikan dampak positif maupun negatif bagi penggunaannya (Rachman 2015). Penggunaan media sosial secara berlebihan akan berdampak buruk bagi seseorang, baik secara fisik maupun secara mental. Kecanduan media sosial dapat berpengaruh pada kecemasan sosial dan gangguan narsistik. Seseorang yang mempunyai kecenderungan gangguan narsistik akan lebih tertarik dengan hal yang hanya menyangkut dengan kesenangan pribadi. Hal ini juga memberikan pengaruh cukup besar dalam pergaulan sehari-hari dan biasanya tidak memiliki kepedulian terhadap perasaan orang lain (Engkus *et al.* 2017). Penderita gangguan kepribadian narsistik yang tidak segera ditangani, akan berdampak negatif pada kehidupan individu dan hubungan si penderita dengan orang lain, bahkan beresiko pada perilaku bunuh diri (Dawood *et al.* 2017). Oleh karena itu, perlu dilakukan diagnosa sejak dini bagi seseorang yang rentan terkena narsistik (Engkus *et al.* 2017).

Jumlah penderita gangguan kepribadian tidak sebanding dengan jumlah tenaga profesional bidang kesehatan mental (psikiater) di Indonesia. Jumlah tenaga profesional bidang kesehatan mental (psikiater) hanya berjumlah 1.053 orang. Hal ini menunjukkan bahwa satu orang psikiater melayani sekitar 250 ribu penduduk (Rokom 2021), sehingga menyebabkan banyaknya penderita gangguan kepribadian yang tidak tertangani. Keterbatasan jumlah psikiater akan menimbulkan masalah dalam hal kecepatan dalam melakukan diagnosa, karena setiap pasien harus menunggu jadwal antrean untuk berkonsultasi sekaligus melakukan terapi pengobatan (Noor 2019). Selain itu, cara yang digunakan untuk mendiagnosis pasien masih menggunakan cara tradisional, proses diagnosa hanya berdasarkan laporan verbal pasien dan sesi wawancara yang panjang dengan psikolog/psikiater. Cara tersebut dapat menghasilkan diagnosa yang bersifat subjektif, mahal, dan membutuhkan waktu yang lama (Jaiswal *et al.* 2019). Kecerdasan buatan dan *machine learning* (ML) diperlukan untuk memperdalam pemahaman tentang situasi kesehatan mental. Kecerdasan buatan dan *machine learning* merupakan metode yang menjanjikan untuk mendukung psikiater dalam membuat hasil diagnosa menjadi lebih baik (Jan *et al.* 2021).

*Machine learning* merupakan salah satu metode kecerdasan buatan yang memungkinkan komputer untuk belajar dan meningkatkan kinerjanya dalam melakukan tugas tertentu secara otomatis, tanpa perlu diprogram secara eksplisit (Reddy *et al.* 2018). *Machine learning* didefinisikan sebagai metode komputasi berdasarkan pengalaman untuk meningkatkan kinerja atau membuat prediksi yang akurat. Berdasarkan data Scopus ([www.scopus.com](http://www.scopus.com)), selama tahun 2016 sampai 2018 jumlah publikasi di bidang kesehatan yang menerapkan *machine learning* meningkat dari 1.658 menjadi 3.904 (pencarian berdasarkan judul, abstrak atau kata kunci) (Fazel dan O'Reilly 2019). Selain itu, penerapan *machine learning* dapat digunakan sebagai model pendekatan untuk mengurangi stress (Reddy *et al.* 2018). Teknik *machine learning* yang paling populer yaitu prediksi, klasifikasi untuk mengidentifikasi ciri-ciri kepribadian (Haz *et al.* 2022). Berdasarkan hasil penelusuran literature, telah banyak penelitian terdahulu yang membahas tentang diagnosa gangguan kepribadian dengan menggunakan teknik *machine learning*.

Penelitian yang dilakukan oleh (Tyagi *et al.* 2018) mengimplementasikan beberapa metode *machine learning* (*Linear Discriminant Analysis* (LDA), *Naive Bayes* (NB), *Classification and Regression Trees* (CART), *K-Nearest-Neighbor* (KNN), *Linear Regression* (LR), dan *Support Vector Machine* (SVM)) untuk mendiagnosa *autism spectrum disorder* (ASD). (Jain *et al.* 2021) menerapkan metode (*Decision tree* (DT), *Random Forest* (RF), SVM, NB, *Logistic Regression* (LR), *XGBoost* (XGB), *Gradient Boosting Classifier* (GBC) dan *Artificial Neural Network* (ANN)). Dari penelitian tersebut, menunjukkan bahwa metode SVM memiliki kinerja yang paling baik yaitu memiliki nilai akurasi sebesar 87,38%. Penelitian selanjutnya oleh (Kim *et al.* 2021) yang melakukan prediksi resiko bunuh diri menggunakan metode *Random Forest* dan *K-Nearest Neighbour* berdasarkan data *Minnesota Multiphasic Personality Inventory-2* (MMPI-2). Hasil penelitian tersebut, menunjukkan bahwa metode *Random Forest* memiliki kinerja lebih baik dibandingkan *K-Nearest Neighbour*, yaitu memiliki nilai akurasi sebesar 92,9% untuk grup *suicidal ideation* dan nilai akurasi sebesar 95% untuk grup *suicidal attempts* (Kim *et al.* 2021).

Berdasarkan hasil penelitian terdahulu tentang penerapan metode *machine learning* untuk memprediksi maupun mendiagnosa gangguan kepribadian telah memberikan kontribusi yang besar pada bidang kesehatan, terutama bidang psikologi. Penelitian yang menerapkan metode *machine learning* pada bidang kesehatan mampu meningkatkan kemampuan dalam proses pengambilan keputusan dengan tingkat keakuratan yang tinggi (Godara dan Singh 2016). Namun sejauh ini, belum ada yang melakukan kajian melalui studi literatur tentang penerapan metode *machine learning* untuk diagnosa gangguan kepribadian. Selain itu, sejauh ini juga belum pernah dilakukan penelitian untuk mencari subset fitur terbaik guna menghasilkan kinerja yang optimal dengan menerapkan teknik pemilihan fitur pada kasus *narcissistic personality disorder* serta belum pernah dilakukan pengukuran dan perbandingan kinerja beberapa metode *machine learning*. Sehingga dalam penelitian akan melakukan: a) literature review dan analisis bibliometrik dengan teknik *systematic literature review* (SLR); b) mencari subset fitur dengan menggunakan teknik *information gain* dan *gain ratio*; c) mengukur dan mengevaluasi kinerja *machine learning* (*Support Vector Machine* (SVM), *Random*



*Forest Classifier (RFC)* dan *Naive Bayes*); d) mengatasi ketidak seimbangan kelas pada data *narcisistic personality disorder* dengan metode *synthetic minority oversampling technique (SMOTE)*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian yang telah dibahas sebelumnya, maka dapat dirumuskan permasalahan penelitian sebagai berikut:

1. Bagaimana proses penelusuran studi literatur dan proses analisis bibliometrik tentang riset kecerdasan buatan (*artificial intelligence: AI*) dalam memprediksi gangguan kepribadian?
2. Bagaimana menentukan subset fitur yang yang paling terkait dengan label kelas guna meningkatkan kinerja metode *machine learning* dalam menganalisis prediksi *Narcisistic Personality Disorder*?
3. Bagaimana mengevaluasi, menganalisis dan melakukan perbandingan kinerja metode *machine learning* dalam prediksi *Narcisistic Personality Disorder*?
4. Bagaimana mengatasi ketidakseimbangan kelas pada metode *machine learning* untuk memprediksi *Narcisistic Personality Disorder*?

## 1.3 Tujuan Penelitian

Berdasarkan permasalahan yang telah diuraikan pada latar belakang, maka tujuan yang dicapai pada penelitian ini sebagai berikut :

1. Melakukan proses penelusuran studi literatur dan analisis bibliometrik tren kecerdasan buatan (*artificial intelligence: AI*) untuk memprediksi gangguan kepribadian dengan menggunakan teknik *systematic literarture review (SLR)* dengan visualisasi menggunakan VOSviewer.
2. Mengevaluasi teknik pemilihan fitur (*feature selection*) guna menghasilkan fitur-fitur yang relevan dan dapat meningkatkan kinerja metode *machine learning*.

3. Mengevaluasi dan melakukan perbandingan kinerja berbagai metode *machine learning* dalam memprediksi *narcisistic personality disorder*.
4. Membangun model *synthetic minority oversampling technique* (SMOTE) untuk mengatasi ketidakseimbangan kelas.

#### 1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini antara lain:

1. Merekomendasikan daftar penelitian yang telah dilakukan dalam penerapan teknik kecerdasan buatan untuk prediksi gangguan kepribadian.
2. Meningkatkan kinerja metode *machine learning* melalui teknik pemilihan fitur.
3. Memberikan rekomendasi pada psikolog dalam memprediksi *Narcisistic Personality Disorder*.
4. Merekomendasikan metode *machine learning* yang memiliki kinerja paling baik dalam prediksi *Narcisistic Personality Disorder*.

#### 1.5 Batasan Masalah

Agar penelitian ini fokus pada permasalahan, maka perlu dilakukan batasan permasalahan. Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Batasan usia yang digunakan pada penelitian ini yaitu 14-50 tahun. Hal ini dikarenakan rentang usia tersebut adalah usia yang rentan terhadap gangguan kepribadian.
2. Dataset yang digunakan dalam penelitian ini diperoleh secara *online* ([www.kaggle.com](http://www.kaggle.com)).
3. Durasi tahun terbit artikel yang digunakan pada analisis bibliometrik yaitu 2018 – 2023.
4. Teknik pemilihan fitur yang diterapkan pada penelitian ini guna menghasilkan subset fitur yang mampu meningkatkan kinerja *machine learning* yaitu Teknik *Gain Ratio* (GR) dan *Information Gain* (IG).

5. Untuk mengatasi ketidakseimbangan kelas pada metode *machine learning* digunakan metode *synthetic minority oversampling technique* (SMOTE).

## 1.6 Kontribusi Penelitian

Kontribusi yang diberikan dari hasil penelitian ini antara lain:

1. Memberikan gambaran tentang tren penelitian dalam menerapkan pendekatan kecerdasan buatan untuk memprediksi dan mendiagnosa gangguan kepribadian. Artikel yang digunakan dalam penelitian ini adalah artikel berbahasa Inggris yang tersedia pada *database Google Scholar*.
2. Mengembangkan model untuk mencari subset fitur yang mampu meningkatkan kinerja metode *machine learning* menggunakan teknik pemilihan fitur *information gain* dan *gain ratio*.
3. Mengevaluasi kinerja berbagai pendekatan *machine learning* guna memperoleh informasi tentang metode terbaik dan waktu pemodelan tercepat dalam memprediksi *narcisistic personality disorder*.
4. Mengembangkan model untuk mengatasi ketidakseimbangan kelas dalam melakukan prediksi *narcisistic personality disorder* agar nilai akurasi mencapai lebih dari 90%.

## 1.7 Keterbaruan (*Novelty*)

Penelitian ini menghasilkan keterbaruan (*novelty*) berupa:

1. Belum ditemukan penelitian yang melakukan studi literatur tentang pendekatan kecerdasan buatan dalam memprediksi gangguan kepribadian. Maka dari itu, dalam penelitian ini melakukan analisis bibliometrik dengan visualisasi menggunakan VOSviewer untuk memetakan penelitian yang terkait dengan pendekatan kecerdasan buatan dalam memprediksi gangguan kepribadian. Teknik yang digunakan berdasarkan kata kunci yang telah ditentukan oleh peneliti dan menggunakan artikel berbahasa Inggris yang tersedia di *database Google Scholar* (artikel yang terindeks *Scopus Quartile Q1-Q4*).

2. Belum ditemukan penelitian yang mengembangkan model teknik pemilihan fitur dan mengevaluasinya untuk memperoleh fitur-fitur yang relevan dengan kelas sehingga mampu meningkatkan akurasi metode *machine learning*. Maka dari itu, penelitian ini melakukan pengembangan model dan mengevaluasi teknik pemilihan fitur dalam memprediksi gangguan kepribadian. Diharapkan dengan dilakukannya penelitian ini, akan menghasilkan subset fitur yang berpengaruh terhadap peningkatan akurasi.
3. Penelitian dengan menerapkan berbagai metode *machine learning* untuk diganosa dan prediksi gangguan kepribadian telah banyak dilakukan, namun yang belum ditemukan penelitian yang membahas tentang pengukuran dan perbandingan kinerja metode *machine learning* untuk diagnosa *narcissistic personality disorder*.
4. Belum ditemukan penelitian yang mengimplementasikan model *synthetic minority oversampling technique* (SMOTE) untuk mengatasi ketidakseimbangan kelas pada *narcisistic personality disorder*. Kondisi ketidakseimbangan kelas menjadi masalah dalam klasifikasi karena *classifier learning* akan condong memprediksi ke kelas mayoritas dibanding dengan kelas minoritas. Akibatnya, dihasilkan akurasi prediksi yang baik terhadap kelas-kelas mayoritas pada data *training* sedangkan untuk kelas-kelas minoritas akan dihasilkan akurasi prediksi yang buruk. Oleh karena itu, penelitian ini perlu dilakukan untuk memberikan temuan baru dalam bidang *machine learning* pada bidang kesehatan berupa model untuk menangani ketidakseimbangan kelas pada *narcisistic personality disorder*.

## II. TINJAUAN PUSTAKA

Pada bagian ini dipaparkan mengenai teori-teori yang mendukung dalam penelitian tentang gangguan mental, kecerdasan buatan dan mesin pembelajaran.

### 2.1. Studi Literatur

Penelitian yang membahas tentang diagnosa gangguan kepribadian menggunakan teknik kecerdasan buatan masih terbatas. Sehingga penelitian dibidang ini perlu dilakukan pengkajian lebih lanjut. Penelitian-penelitian tersebut dapat dilihat pada tabel 1.

Tabel 1. Penelitian terdahulu

Penulis/Tahun	Metode	Jenis Gangguan Kepribadian	Data
(Adji <i>et al.</i> 2015)	Sistem Pakar dengan Metode <i>Certainty Factor</i>	Paranoid, Anti-Sosial, Histrionik, Obsesif, Kecemasan dan Dependen	10 pasien yang mengalami gangguan kepribadian yang diobservasi oleh 2 psikolog. Untuk mendeskripsikan jenis gangguan, digunakan data kode gejala.
(Laijawala <i>et al.</i> 2020)	<i>Decision Tree, Random Forest, Naive Bayes</i>	Penyakit Mental	Data <i>online</i> yang bersumber dari OSMI ( <i>Open Sourcing Mental Illness</i> ) survey.
(Srividya <i>et al.</i> 2018a)	<i>Logistic Regression, Naive Bayes, SVM, Decision Tree, KNN, Ensemble (Bagging), Random Forest</i>	Penyakit Mental	Menggunakan 656 data sampel dengan jumlah fitur sebanyak 20 dan 3 kelas. Usia rata-rata yang menjadi sampel yaitu 22 tahun.

Penulis/Tahun	Metode	Jenis Gangguan Kepribadian	Data
(Al-hajji <i>et al.</i> 2019)	<i>Rule-Based Expert System (RBES)</i>	<i>Psychological anxiety, Obsessive-compulsive disorder, hysteria, depresi</i>	23 gejala gangguan kejiwaan.
(Huang <i>et al.</i> 2018)	<i>Convolutional Neural Network dan Long Short-term Memory</i>	<i>Mood Disorder</i>	Data Ungkapan emosional dari 15 orang dengan gangguan bipolar, 15 orang dengan unipolar, dan 15 data normal/sehat.
(Bilek <i>et al.</i> 2019)	<i>Neural Network</i>	Gangguan Depresi Mayor	<i>Functional magnetic resonance imaging (fMRI)</i> dari 37 pasien MDD ( <i>Mayor Depressive Disorder</i> ) dan 33 data <i>check-up</i> sehat.
(Schlumpf <i>et al.</i> 2019)	<i>Neural Network</i>	<i>Complex Trauma Disorder</i>	Hasil pengisian kuesioner pasien dan uji coba EEG.
(Janssen <i>et al.</i> 2014)	<i>Case-Based Reasoning</i>	Gangguan mental, kecuali gangguan kepribadian dan Retadasi Mental	Menggunakan 30 fitur dari hasil pre test.
(Jung dan Yoon 2017)	<i>Fuzzy Logic</i>	Tingkat stress	Menggunakan multimodal biosensor, tekanan darah, detak jantung (ECG), tingkat pernafasan, and detak jantung yang diukur berulang kali dan data dikumpulkan dari 4 orang pasien.
(Rachakonda dan Kougianos 2018)	<i>Fuzzy Logic</i>	Tingkat stress	Suhu, aktivitas fisik dan data kelembapan yang diperoleh dari sensor.
(Silvana <i>et al.</i> 2018)	<i>Fuzzy Logic</i>	Gangguan Mental	Data dari pengisian kuesioner yang didistribusikan kepada 16 responden pilihan.

Penulis/Tahun	Metode	Jenis Gangguan Kepribadian	Data
			Data berisikan daftar gejala dan tipe gangguan mental.

## 2.2. Gangguan Mental (*Mental Disorder*)

Seseorang yang mengalami kegagalan dalam menyesuaikan diri atau menyelesaikan masalah internal dan eksternal dapat menjadi penyebab adanya gangguan mental (Konna 2017). Kesehatan mental didefinisikan sebagai ketika seseorang percaya pada kemampuan mereka sendiri, dapat mengatasi tantangan sehari-hari, bekerja secara produktif, dan mampu memberikan kontribusi kepada komunitasnya (Ayuningtyas 2017). Gangguan mental dapat menimbulkan stres bagi penderita dan keluarganya. Gangguan mental dapat terjadi pada setiap orang, tanpa mengenal umur, ras, agama, maupun status sosial-ekonomi.

Gangguan mental, dikenal juga sebagai penyakit mental atau gangguan psikologis, merujuk pada kondisi yang mempengaruhi pola pikir, suasana hati, dan perilaku seseorang. Gangguan mental dapat memengaruhi berbagai aspek kehidupan seseorang, termasuk hubungan interpersonal, fungsi sosial, kinerja kerja, dan kualitas hidup secara keseluruhan. Ada banyak jenis gangguan mental yang berbeda dengan gejala, penyebab, dan perawatan yang beragam.

Kesehatan mental juga pernah menjadi trending *google* bersamaan dengan Hari Bumi pada 22 April 2022 lalu. Hal tersebut menunjukkan bahwa masalah kesehatan mental di Indonesia masih menjadi persoalan serius. Berdasarkan data Riset Kesehatan Dasar (Riskesdas) 2018 ((KemenkesRI) 2019), lebih dari 19 juta penduduk berusia lebih dari 15 tahun mengalami gangguan mental emosional, dan lebih dari 12 juta penduduk berusia lebih dari 15 tahun mengalami depresi. Gejala permulaan bagi seseorang yang mengalami gangguan mental antara lain (Kessler, *et al.*, 2005):

- 1) tampak fisik dengan gejala seperti pusing, sesak napas, demam, dan nyeri lambung;
- 2) menunjukkan perasaan cemas, ketakutan, patah hati, apatis, cemburu, dan kemarahan dalam dirinya.

Sedangkan sebab timbulnya gangguan mental antara lain (Kessler *et al.* 2005):

- 1) kepribadian yang lemah akibat kondisi jasmani atau mental yang kurang sempurna;
- 2) terjadinya konflik sosial budaya;
- 3) cara pematangan batin yang salah dengan memberikan reaksi yang berlebihan terhadap kehidupan sosial.

Proses gangguan mental yang dialami seseorang mendorongnya kearah positif dan negatif. Positif dapat berupa trauma jiwa yang dialami diatasi dengan baik, sebagai usaha agar tetap *survive* dalam hidup, misalnya melakukan sholat tahajud, ataupun melakukan kegiatan yang positif. Sedangkan arah negatif dapat berupa trauma yang dialami diperlarutkan sehingga yang bersangkutan mengalami frustasi, yaitu tekanan batin akibat tidak tercapainya apa yang diinginkan. Penyebab lain munculnya gangguan mental pada seseorang bisa terjadi karena tidak terpenuhinya kebutuhan (Mubasyaroh 2013).

Penyebab gangguan mental sangat kompleks dan melibatkan kombinasi faktor genetik, lingkungan, biologis, dan psikososial. Faktor-faktor seperti keturunan, trauma, stres, ketidakseimbangan kimia otak, dan gangguan perkembangan dapat memainkan peran dalam munculnya gangguan mental. Perawatan gangguan mental dapat melibatkan pendekatan yang berbeda, termasuk terapi kognitif perilaku (CBT), terapi obat, terapi bicara, terapi kelompok, dan intervensi lainnya. Pilihan perawatan tergantung pada jenis dan tingkat keparahan gangguan mental, serta preferensi individu. Berikut adalah beberapa contoh gangguan mental yang umum (Kessler *et al.* 2005):



a) Gangguan Kecemasan

Gangguan kecemasan mencakup beberapa jenis gangguan antara lain gangguan kecemasan umum, gangguan panik, gangguan stres pasca trauma, fobia, dan gangguan kecemasan sosial. Gejala dari gangguan kecemasan meliputi kecemasan yang berlebihan, ketegangan, kegelisahan, serangan panik, dan perasaan takut yang intens.

b) Gangguan Mood

Gangguan mood mencakup depresi, bipolar, dan gangguan suasana hati lainnya. Depresi ditandai dengan perasaan sedih yang berkepanjangan, kehilangan minat, perubahan berat badan, perubahan energi, dan pikiran tentang bunuh diri. Gangguan bipolar melibatkan perubahan suasana hati ekstrem antara depresi dan mania.

c) Gangguan Makan

Beberapa jenis gangguan makan antara lain anoreksia nervosa, bulimia nervosa, dan gangguan makan lainnya. Gangguan makan ditandai dengan pola makan yang tidak sehat, obsesi dengan berat badan, kecemasan tentang penampilan fisik, dan dampak negatif pada kesehatan fisik dan emosional.

d) Gangguan Perilaku

Gangguan perilaku meliputi beberapa jenis gangguan, antara lain gangguan defisit perhatian dan hiperaktivitas (ADHD), gangguan perilaku eksplosif, gangguan kondukt, dan gangguan perilaku lainnya. Gejalanya termasuk impulsivitas, hiperaktivitas, ketidakmampuan untuk berkonsentrasi, dan perilaku yang bertentangan dengan norma sosial.

e) Gangguan Psikotik

Jenis gangguan psikotik yaitu seperti skizofrenia dan gangguan psikotik lainnya. Gejala psikotik meliputi delusi (keyakinan yang salah), halusinasi (persepsi yang tidak ada), dan gangguan pikiran yang parah.

f) Gangguan Kepribadian

Jenis-jenis gangguan kepribadian antara lain gangguan kepribadian *borderline*, antisosial, *narcisistic*, dan lainnya. Gangguan kepribadian ditandai dengan pola perilaku yang persisten, tidak sehat, dan bermasalah dalam hubungan interpersonal.

Salah satu gangguan mental yang sedang populer adalah gangguan kepribadian. Gangguan kepribadian adalah kategori gangguan mental yang ditandai oleh pola perilaku yang tidak sehat, maladaptif, dan persisten yang mempengaruhi cara individu berpikir, merasa, dan berinteraksi dengan orang lain. Gangguan kepribadian memengaruhi pola pikir, emosi, persepsi, dan perilaku individu, dan cenderung berlangsung dalam jangka waktu yang panjang. Ada beberapa jenis gangguan kepribadian yang diakui dalam *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), yaitu manual diagnostik yang digunakan oleh para profesional kesehatan mental. Beberapa contoh gangguan kepribadian yang umum meliputi (Dilip V. Jeste *et al.* 2013):

a) Gangguan Kepribadian *Borderline* (*Borderline Personality Disorder*/BPD)

Jenis gangguan kepribadian ini ditandai dengan adanya ketidakstabilan emosional yang ekstrem, impulsivitas, pola hubungan yang tidak stabil, dan ketakutan akan penolakan atau keabandonan. Individu dengan BPD sering mengalami perubahan mood yang drastis, masalah pengendalian diri, dan kesulitan dalam mempertahankan hubungan yang stabil.

b) Gangguan Kepribadian Antisosial (*Antisocial Personality Disorder*/ASPD)

Jenis gangguan kepribadian ini dicirikan oleh perilaku yang tidak peduli terhadap norma sosial, penipuan, manipulasi, dan ketidakmampuan untuk memperhatikan atau memahami perasaan dan kebutuhan orang lain. Individu dengan ASPD cenderung melanggar hak-hak orang lain, seringkali terlibat dalam perilaku kriminal, dan kurang memiliki empati atau penyesalan atas tindakan yang telah dilakukan.

c) Gangguan Kepribadian *Narcissistic* (*Narcissistic Personality Disorder*/NPD)

Jenis gangguan kepribadian ini ditandai oleh kebutuhan yang berlebihan akan perhatian dan pengakuan, perasaan superioritas, kurangnya empati terhadap orang lain, dan kecenderungan untuk memanipulasi orang lain untuk memenuhi kebutuhannya sendiri. Individu dengan *narcissistic personality disorder* sering memiliki harga diri yang terlalu tinggi dan mencari pengakuan yang berlebihan.

d) Gangguan Kepribadian Obsesif-Kompulsif (*Obsessive-Compulsive Personality Disorder/OCPD*)

Jenis gangguan kepribadian ini dicirikan oleh keterlibatan yang berlebihan dengan ketertiban, kesempurnaan, dan kontrol. Individu dengan OCPD memiliki pola pikir yang kaku, perfeksionis, dan sulit menerima ketidakpastian. Penderita cenderung sangat fokus pada rincian dan aturan, dan kesulitan dalam beradaptasi dengan perubahan.

e) Gangguan Kepribadian Paranoid (*Paranoid Personality Disorder/PPD*)

Jenis gangguan kepribadian ini ditandai oleh kecurigaan yang ekstrem, ketidakpercayaan terhadap orang lain, dan interpretasi yang negatif terhadap motif dan niat orang lain. Individu dengan PPD cenderung waspada dan merasa terancam secara terus-menerus.

Saat ini, seiring dengan berkembangnya teknologi berpengaruh pada peningkatan pengguna media sosial. Hal tersebut berdampak pada maraknya orang yang terkena *narcissistic personality disorder* tanpa disadari. Dewasa ini internet dan media sosial adalah alat bagi individu gangguan kepribadian narsistik untuk mengaktualkan dirinya sendiri, membesar-besarkan dirinya dengan memposting foto atau video prestasi dan berbagai potensi ke media sosial dengan harapan mendapatkan pengakuan dan apresiasi dari orang lain (Santi 2017). *Narcissistic personality disorder* dapat memiliki dampak yang signifikan pada hubungan interpersonal, pekerjaan, dan kualitas hidup secara keseluruhan. Orang dengan *narcissistic personality disorder* cenderung mengalami kesulitan dalam mempertahankan hubungan yang stabil, karena penderita berfokus yang berlebihan pada diri sendiri dan kurangnya empati terhadap orang lain. Berikut adalah beberapa ciri khas *narcissistic personality disorder* (Dilip V. Jeste *et al.* 2013):

a) Rasa Kepentingan yang Berlebihan

Seseorang yang terdiagnosa *narcissistic personality disorder* cenderung memiliki rasa kepentingan yang berlebihan terhadap diri sendiri. Penderita sering merasa bahwa dirinya istimewa, unik, dan berhak mendapatkan perlakuan istimewa.

b) Fantasi Kehebatan dan Kesuksesan

Seseorang yang terdiagnosa *narcissistic personality disorder* sering memiliki fantasi yang berlebihan tentang kehebatan, kecerdasan, atau kekuasaan. Penderita cenderung berbicara tentang pencapaian yang luar biasa tanpa memiliki bukti yang substansial.

c) Perlu Diperhatikan dan Dipuji

Seseorang yang terdiagnosa *narcissistic personality disorder* sangat membutuhkan perhatian dan pengakuan dari orang lain. Penderita ingin menjadi pusat perhatian di setiap situasi dan sering kali memanipulasi orang lain untuk memperoleh perhatian dan pujian.

d) Kurangnya Empati

Seseorang yang terdiagnosa *narcissistic personality disorder* cenderung memiliki kurangnya empati terhadap perasaan dan kebutuhan orang lain. Penderita merasa kesulitan dalam memahami dan merasakan emosi orang lain, dan seringkali kurang peduli dengan perasaan dan perspektif orang lain.

e) Perilaku Manipulatif

Seseorang yang terdiagnosa *narcissistic personality disorder* sering menggunakan manipulasi dan eksploitasi terhadap orang lain untuk mencapai tujuan. Penderita cenderung memanipulasi orang lain untuk memperoleh keuntungan pribadi dan menggunakan orang lain sebagai alat untuk memperkuat harga dirinya.

f) Reaksi yang Terlalu Sensitif terhadap Kritik

Individu dengan *narcissistic personality disorder* sering marah secara berlebihan ketika menerima kritik atau penghinaan. Penderita tidak mampu menerima kritik dengan bijaksana dan merasa terancam oleh pendapat atau pandangan negatif tentang dirinya.

### 2.3. Kecerdasan Buatan (*Artificial Intelligence*)

Memprediksi gangguan kepribadian menggunakan *Artificial Intelligence (AI)* menjadi salah satu bidang penelitian yang baru dan akan menjadi topik pembahasan di masa yang akan datang. Kecerdasan buatan telah mendapatkan perhatian yang

luar biasa dalam beberapa tahun terakhir dari akademisi, industri dan masyarakat umum. Bagian utama dari aplikasi kecerdasan buatan adalah adanya pengetahuan yang merupakan suatu pengertian tentang beberapa wilayah subyek yang diperoleh melalui pendidikan dan pengalaman. Adanya pengetahuan dan kemampuan untuk menarik kesimpulan dari pengalaman, komputer dapat dijadikan alat bantu yang praktis untuk memecahkan masalah dan pengambilan keputusan.

Teknologi kecerdasan buatan tidak lagi menjadi ranah futurologis tetapi merupakan komponen integral dari model bisnis organisasi dan elemen strategis utama dalam rencana sektor bisnis, kedokteran, dan pemerintah dalam skala global (Dwivedi *et al.* 2019). Kecerdasan buatan memiliki berbagai pendekatan yang berbeda, mulai dari representasi pengetahuan *top-down* hingga *machine learning bottom-up* (Hu *et al.* 2019). Teknik kecerdasan buatan dan pemodelan matematis semakin banyak diperkenalkan untuk membantu memecahkan permasalahan dalam penelitian kesehatan mental (Liu *et al.* 2020). Teknik kecerdasan buatan ini mampu memperoleh informasi secara rinci untuk mengetahui ciri-ciri dalam memprediksi gangguan kepribadian (Jan *et al.* 2018). Seiring dengan meningkatnya kemampuan prosesor komputer, berbagai metode berbasis kecerdasan buatan telah dikembangkan sebagai pendekatan untuk memodelkan masalah yang kompleks (Seyedzadeh *et al.* 2019). Teknik kecerdasan buatan telah banyak digunakan untuk proses pengambilan keputusan. Tujuan utama dari pengembangan kecerdasan buatan adalah menciptakan mesin yang dapat memahami, belajar, merencanakan, dan menyelesaikan masalah dengan cara yang mirip dengan manusia. Konsep kecerdasan buatan melibatkan pengembangan dan penerapan teknik-teknik komputasi yang memungkinkan mesin untuk menunjukkan kemampuan seperti persepsi, pemahaman bahasa, pembelajaran, penalaran, pemecahan masalah, dan pengambilan keputusan.

Beberapa keuntungan dan kelemahan dalam penerapan teknik kecerdasan buatan antara lain:

1. Penggunaan komputer dimasa yang akan datang dapat memberikan kenyamanan dan kesenangan bagi penggunanya, namun hal tersebut akan

memicu pada harga komputer yang akan menjadi semakin mahal. Para pengguna komputer akan tertarik dengan beberapa hal yang ditawarkan oleh adanya kecerdasan buatan, antara lain:

- a) mampu berkomunikasi dengan komputer menggunakan bahasa alami/bahasa manusia sehari-hari;
  - b) terbebas dari kewajiban dalam belajar bahasa pemrograman dan sistem operasi;
  - c) dengan menggunakan komputer, para pengguna komputer yang tidak terlatih sekalipun akan menghasilkan karya yang sangat berguna bagi kepentingannya;
  - d) penggunaan komputer akan tidak lebih sulit daripada penggunaan perangkat telepon/seluler.
2. Komputer akan menjadi semakin lebih berguna, melalui kemampuannya dalam menyelesaikan permasalahan-permasalahan menggunakan teknik AI.
  3. Pengembangan dan penelitian kecerdasan buatan membutuhkan biaya yang sangat mahal.
  4. Pengembangan aplikasi kecerdasan buatan membutuhkan waktu yang sangat lama.
  5. Proses pembuatan dan pengembangan perangkat lunak AI dapat cepat selesai dengan menggunakan perangkat khusus AI, namun belum terlalu banyak yang membuatnya. Penggunaan perangkat lunak khusus dapat mempercepat dan mempermudah proses pembuatan dan pengembangan perangkat lunak kecerdasan buatan.

Aplikasi kecerdasan buatan terdiri dari dua bagian utama, yaitu:

1. Basis pengetahuan (*knowledge based*) yang berisi fakta tentang objek-objek dalam domain yang dipilih dan hubungan diantara domain-domain tersebut.
2. Motor Inferensi (*inference engine*) merupakan sekumpulan cara yang digunakan untuk menguji basis pengetahuan dalam menjawab suatu pertanyaan, menyelesaikan masalah, atau membuat keputusan. *Inference engine* memiliki kemampuan dalam menarik kesimpulan berdasarkan pengalaman.

Beberapa bidang kecerdasan buatan yang saat ini banyak dikembangkan, antara lain:

1) **Sistem Pakar (*expert system*)** merupakan sistem komputer yang dirancang untuk meniru pengetahuan dan keterampilan seorang pakar manusia di bidang tertentu. Tujuan utama dari sistem pakar adalah memberikan solusi yang cerdas dan tepat dalam menghadapi masalah yang kompleks, dengan cara menerapkan pengetahuan dan aturan yang diperoleh dari para pakar manusia. Sistem pakar memiliki konsep dasar dalam pengembangannya, antara lain:

a) Basis Pengetahuan

Sistem pakar memanfaatkan pengetahuan yang dikumpulkan dari pakar manusia dalam bidang yang spesifik. Pengetahuan ini biasanya bersifat heuristik dan berdasarkan pengalaman dan pemahaman mendalam para pakar. Pengetahuan pakar dapat dinyatakan dalam bentuk aturan, kasus-kasus, atau model matematika.

b) Penalaran

Sistem pakar menggunakan penalaran untuk mengaplikasikan pengetahuan pakar terhadap masalah yang dihadapi. Penalaran ini dilakukan dengan menerapkan aturan-aturan dan algoritma yang ada dalam sistem pakar. Berdasarkan informasi yang diberikan, sistem pakar dapat mencapai kesimpulan atau memberikan rekomendasi.

c) Representasi Pengetahuan

Pengetahuan pakar dalam sistem pakar dapat direpresentasikan dalam berbagai bentuk, seperti aturan produksi (if-then), jaringan semantik, atau grafik pengetahuan. Representasi ini memungkinkan sistem pakar untuk memahami dan memanfaatkan pengetahuan secara efektif.

d) Sistem Inferensi

Sistem inferensi dalam sistem pakar berfungsi untuk melakukan penalaran berdasarkan pengetahuan yang dimiliki. Sistem inferensi mengambil fakta-fakta dan aturan-aturan yang ada, kemudian menghasilkan kesimpulan atau rekomendasi berdasarkan masalah yang dihadapi.

Penerapan sistem pakar dapat ditemukan di berbagai bidang, antara lain:

a) Kedokteran

Sistem pakar kedokteran dapat membantu dalam diagnosis penyakit dengan memeriksa gejala, riwayat medis, dan hasil tes pasien. Sistem pakar juga dapat memberikan rekomendasi pengobatan berdasarkan informasi yang ada.

b) Keuangan

Sistem pakar keuangan dapat memberikan rekomendasi investasi berdasarkan profil risiko dan tujuan keuangan seseorang. Sistem pakar juga dapat membantu dalam analisis risiko dan manajemen portofolio.

c) Manufaktur

Sistem pakar dapat digunakan dalam pengendalian kualitas, perawatan mesin, perencanaan produksi, dan pemecahan masalah teknis di lingkungan manufaktur.

d) Perancangan dan Teknik

Sistem pakar dapat digunakan dalam perancangan produk, perencanaan layout, analisis struktural, dan peramalan permintaan.

Keuntungan dari penerapan sistem pakar adalah sebagai berikut:

- a) Kemampuan untuk meniru pengetahuan pakar manusia yang terbatas dalam jumlah dan tersebar secara luas.
- b) Konsistensi dalam pengambilan keputusan, di mana sistem pakar akan menghasilkan solusi yang konsisten berdasarkan aturan-aturan yang ditetapkan.
- c) Peningkatan kecepatan dan efisiensi dalam pemecahan masalah atau memberikan rekomendasi.
- d) Dapat berfungsi sebagai alat pelatihan bagi individu yang ingin mempelajari bidang tertentu.

Namun, penerapan sistem pakar juga memiliki beberapa tantangan, seperti kesulitan dalam mengakuisisi pengetahuan pakar dengan akurasi tinggi, perubahan yang cepat dalam pengetahuan domain, dan kesulitan dalam memodelkan pengetahuan yang bersifat ambigu atau tidak pasti.



- 2) **Pemrosesan Bahasa Alami** (*natural language processing: NLP*) merupakan cabang kecerdasan buatan yang berfokus pada interaksi antara manusia dan komputer menggunakan bahasa manusia alami. *Natural language processing* melibatkan pemrosesan, analisis, dan generasi bahasa manusia alami, sehingga komputer dapat memahami, menganalisis, dan menghasilkan teks yang dimengerti oleh manusia.

Keunggulan dari penerapan *natural language processing* antara lain:

- a) Pemrosesan Bahasa Alami: *natural language processing* memungkinkan komputer untuk memahami bahasa manusia alami, yang memungkinkan interaksi yang lebih intuitif antara manusia dan mesin. Hal ini memungkinkan pengguna untuk berkomunikasi dengan sistem menggunakan bahasa sehari-hari tanpa perlu menguasai bahasa pemrograman atau antarmuka teknis.
- b) Analisis Skala Besar: *natural language processing* memungkinkan analisis teks dalam jumlah besar dengan cepat dan efisien. Ini memungkinkan pengolahan dan analisis data yang melibatkan jumlah teks yang besar, seperti ulasan produk, posting media sosial, artikel berita, dan dokumen bisnis.
- c) Klasifikasi dan Kategorisasi Otomatis: *natural language processing* dapat digunakan untuk mengklasifikasikan dan mengkategorikan teks dalam kategori yang relevan. Ini dapat membantu dalam penyaringan informasi, pengindeksan konten, dan pengorganisasian dokumen.
- d) Penerjemahan Mesin: *natural language processing* memungkinkan penerjemahan otomatis antara bahasa yang berbeda. Ini memfasilitasi komunikasi lintas bahasa dan membuka peluang kolaborasi global yang lebih besar.
- e) Analisis Sentimen: *natural language processing* dapat digunakan untuk menganalisis sentimen atau opini dalam teks, seperti ulasan produk atau tanggapan media sosial. Hal ini membantu dalam memahami persepsi dan pendapat pengguna terhadap suatu produk, merek, atau topik tertentu.

Sedangkan kelemahan dari penerapan *natural language processing* antara lain:

- a) Kompleksitas Bahasa: Bahasa manusia sangat kompleks dan ambigu. Pemahaman konteks, makna ganda, atau perubahan makna kata dapat menjadi tantangan dalam *natural language processing*. Keterbatasan dalam pemahaman konteks dapat menyebabkan kesalahan atau interpretasi yang salah.
- b) Variasi dan Slang Bahasa: Bahasa manusia memiliki variasi dialek, gaya penulisan, dan slang yang berbeda-beda. Pemrosesan teks dalam bahasa yang tidak baku atau bahasa yang berubah-ubah dapat menjadi sulit dalam *natural language processing*.
- c) Kesalahan Pemahaman: Pemrosesan bahasa manusia seringkali melibatkan ambiguitas dan perbedaan interpretasi. Meskipun kemajuan dalam *natural language processing*, masih ada kesalahan pemahaman yang dapat terjadi, terutama dalam konteks yang kompleks atau abstrak.

*Natural language processing* telah diterapkan secara luas dalam berbagai bidang dengan tujuan meningkatkan efisiensi, memahami konten teks, dan memberikan pengalaman pengguna yang lebih baik. Berikut ini adalah beberapa contoh penerapan *natural language processing* di berbagai bidang:

- a) Mesin Pencari dan Pemrosesan Informasi

*Natural language processing* digunakan dalam mesin pencari seperti Google untuk memahami pertanyaan pengguna dan memberikan hasil pencarian yang relevan. Pemrosesan bahasa juga digunakan dalam ekstraksi informasi dari dokumen dan penyusunan ringkasan otomatis.

- b) Penerjemahan Mesin

*Natural language processing* digunakan dalam sistem penerjemahan mesin untuk menerjemahkan teks dari satu bahasa ke bahasa lain. Algoritma pemrosesan bahasa alami membantu dalam memahami struktur dan makna teks dalam bahasa sumber, serta menghasilkan teks yang setara dalam bahasa target.

- c) Chatbot dan Asisten Virtual

*Natural language processing* digunakan untuk memungkinkan interaksi yang lebih manusiawi antara pengguna dan chatbot atau asisten virtual. Sistem ini

menggunakan teknik *natural language processing* untuk memahami dan merespons pertanyaan atau perintah dalam bahasa alami, memberikan informasi, menawarkan bantuan, atau melakukan tugas tertentu.

d) Analisis Sentimen

*Natural language processing* digunakan untuk menganalisis sentimen atau opini dalam teks, seperti ulasan produk, tanggapan media sosial, atau survei pelanggan. Analisis sentimen membantu perusahaan dalam memahami sikap dan pendapat pelanggan terhadap produk atau layanan, sehingga dapat mengambil tindakan yang tepat.

e) Pengolahan Bahasa untuk Media Sosial

*Natural language processing* digunakan untuk menganalisis dan memahami konten di platform media sosial. Hal ini dapat mencakup pemrosesan teks dalam tweet, posting Facebook, atau komentar di Instagram untuk mengidentifikasi tren, topik populer, atau pemahaman terhadap preferensi pengguna.

f) Pemrosesan Bahasa untuk Kedokteran

NLP digunakan dalam pemrosesan catatan medis, literatur ilmiah, atau data kesehatan untuk membantu dalam diagnosis, penelitian, dan pengambilan keputusan medis. Penerapan *natural language processing* dalam bidang kesehatan dapat membantu dalam identifikasi gejala, penentuan diagnosis, dan rekomendasi pengobatan.

g) Analisis Teks dan Klasifikasi Dokumen

*Natural language processing* digunakan dalam pemrosesan teks untuk mengklasifikasikan dan mengategorikan dokumen dalam kategori yang relevan. Contohnya adalah pengorganisasian berita berdasarkan topik, pengindeksan konten untuk mesin pencari, atau pemrosesan teks hukum untuk mencari informasi hukum yang relevan.

h) Pengolahan Bahasa untuk E-commerce

*Natural language processing* digunakan dalam industri e-commerce untuk meningkatkan pengalaman pengguna dan personalisasi. Misalnya, sistem merekomendasikan produk berdasarkan preferensi pengguna, memahami

pertanyaan pelanggan, atau melakukan analisis ulasan produk untuk mengidentifikasi fitur yang diinginkan.

i) Pemrosesan Bahasa untuk Keuangan

*Natural language processing* digunakan dalam analisis teks di bidang keuangan untuk mengidentifikasi informasi pasar, mengawasi berita ekonomi, atau menganalisis laporan keuangan perusahaan. Hal ini membantu dalam pengambilan keputusan investasi dan pemodelan risiko.

j) Pengolahan Bahasa untuk Pendidikan

*Natural language processing* digunakan dalam pengembangan sistem pembelajaran adaptif dan evaluasi otomatis. Misalnya, sistem dapat menganalisis respon siswa dalam tes atau tugas, memberikan umpan balik, dan menyesuaikan materi pembelajaran berdasarkan kebutuhan individu.

Penerapan *natural language processing* terus berkembang dan diperluas ke berbagai bidang lainnya, termasuk otomatisasi proses bisnis, analisis hukum, deteksi penipuan, dan lain-lain.

- 3) **Pemahaman Ucapan/Suara (*Speech/Voice Understanding*)** merupakan cabang dari *natural language processing* yang berfokus pada pemrosesan dan pemahaman ucapan manusia. Tujuan utamanya adalah untuk mengubah ucapan manusia ke dalam bentuk teks yang dapat dipahami oleh komputer, serta memahami dan menafsirkan makna yang terkandung dalam ucapan tersebut.

Manfaat dari pemahaman ucapan/suara adalah sebagai berikut:

a) Penginteraksi Manusia-Komputer yang Lebih Alami

Kemampuan memahami ucapan manusia, komputer dapat berinteraksi dengan pengguna secara lebih alami dan intuitif melalui suara. Hal tersebut membantu dalam meningkatkan pengalaman pengguna dan mengurangi hambatan komunikasi.

b) Pengurangan Beban Kerja

Adanya Pemahaman Ucapan, manusia dapat berkomunikasi dengan komputer melalui suara, mengurangi kebutuhan untuk mengetik atau

menggunakan antarmuka lainnya. Hal ini meningkatkan efisiensi dan mengurangi beban kerja pengguna.

c) Asisten Virtual dan Chatbot Suara

Pemahaman Ucapan memungkinkan pengembangan asisten virtual dan chatbot suara yang dapat membantu pengguna dalam menjawab pertanyaan, memberikan informasi, dan melakukan tugas-tugas tertentu melalui interaksi suara.

d) Layanan Pelanggan yang Lebih Baik

Pemahaman Ucapan, perusahaan dapat mengimplementasikan sistem suara yang dapat memahami keluhan atau permintaan pelanggan secara langsung. Hal ini memungkinkan penanganan kasus yang lebih cepat dan peningkatan kepuasan pelanggan.

e) Analisis dan Pemantauan Percakapan

Pemahaman Ucapan juga dapat digunakan dalam analisis dan pemantauan percakapan suara, seperti percakapan layanan pelanggan atau wawancara. Hal ini membantu dalam mengidentifikasi tren, sentimen, atau masalah tertentu dalam percakapan tersebut.

- 4) **Sistem Sensor dan Robotika** merupakan dua bidang yang saling terkait dan sering digunakan bersama-sama dalam pengembangan sistem otomatis. Sistem sensor digunakan untuk mendeteksi, mengukur, dan mengonversi fenomena fisik menjadi sinyal yang dapat diproses oleh komputer. Robotika, di sisi lain, berkaitan dengan desain, pengembangan, dan penggunaan robot untuk melakukan tugas-tugas tertentu.

Sistem sensor terdiri dari berbagai jenis sensor yang digunakan untuk mengumpulkan informasi dari lingkungan fisik. Sensor ini dapat mendeteksi berbagai parameter, seperti suhu, tekanan, cahaya, suara, gerakan, dan banyak lagi. Beberapa contoh sensor yang umum digunakan termasuk sensor suhu, sensor tekanan, sensor optik, sensor gerak, sensor getaran, dan sensor gas.

Manfaat dari sistem sensor adalah sebagai berikut:

a) Pengumpulan Data

Sistem sensor memungkinkan pengumpulan data *real-time* tentang lingkungan fisik. Data ini dapat digunakan untuk analisis, pemantauan, dan pengambilan keputusan yang lebih baik dalam berbagai aplikasi, seperti manufaktur, pertanian, kesehatan, dan lingkungan.

b) Automatisasi Proses

Sensor dapat digunakan untuk mendeteksi perubahan dalam lingkungan dan mengaktifkan tindakan otomatis yang sesuai. Contohnya termasuk pengaturan suhu otomatis, pengendalian kualitas produk, atau pemantauan kebocoran dalam sistem pipa.

c) Keamanan dan Keandalan

Sensor dapat digunakan untuk mendeteksi situasi berbahaya atau potensi kegagalan dalam sistem. Misalnya, sensor asap yang digunakan dalam alarm kebakaran atau sensor kecepatan dalam kendaraan untuk mencegah kecelakaan.

Sedangkan robotika melibatkan penggunaan robot dan teknologi terkait untuk melakukan tugas-tugas tertentu. Robot adalah sistem otomatis yang dapat melakukan tindakan fisik, baik dengan kontrol langsung dari manusia atau dengan menggunakan kecerdasan buatan dan sistem sensor.

Manfaat dari robotika adalah sebagai berikut:

a) Otomatisasi dan Efisiensi

Robotika memungkinkan otomatisasi tugas-tugas yang monoton, berulang, atau berbahaya. Hal ini membantu meningkatkan efisiensi, produktivitas, dan kualitas dalam berbagai sektor, seperti manufaktur, industri, dan pertanian.

b) Peningkatan Keamanan

Robotika digunakan untuk tugas-tugas berbahaya yang berpotensi membahayakan manusia, seperti penyelamatan di tempat yang sulit dijangkau atau penghilangan bahan berbahaya. Robot juga dapat digunakan untuk pengawasan dan pengendalian di lingkungan yang berisiko tinggi.

c) Presisi dan Akurasi

Robotika memungkinkan tugas-tugas yang membutuhkan presisi tinggi dan akurasi, yang sulit dilakukan oleh manusia. Misalnya, robot bedah dapat melakukan prosedur bedah dengan ketepatan yang lebih tinggi dibandingkan dengan tangan manusia.

d) Peningkatan Pengalaman Pengguna

Robotika dapat digunakan dalam aplikasi yang meningkatkan pengalaman pengguna, seperti asisten pribadi atau robot sosial. Robot-robot ini dapat berinteraksi dengan manusia, memahami ucapan atau gerakan, dan memberikan respons yang sesuai.

Penerapan Sistem Sensor dan Robotika dapat ditemukan dalam berbagai bidang, antara lain:

- a) Manufaktur: Penggunaan robot dalam proses produksi untuk perakitan, pengelasan, pengepakan, dan pemrosesan.
- b) Transportasi: Penggunaan sensor dan sistem navigasi pada kendaraan otonom dan robot pengantar.
- c) Kesehatan: Penggunaan robot bedah, perawatan kesehatan jarak jauh, serta sensor untuk pemantauan pasien dan diagnosis medis. Pertanian: Penggunaan robot dan sensor dalam pertanian presisi, pengumpulan data tanah, dan pemantauan tanaman.
- d) Eksplorasi Luar Angkasa: Penggunaan robot untuk eksplorasi luar angkasa, seperti pendaratan pada planet atau asteroid, dan pengumpulan data.
- e) Rumah Pintar: Penggunaan robotika dan sistem sensor dalam rumah pintar untuk mengendalikan suhu, keamanan, dan efisiensi energi.
- f) Layanan Pelanggan: Penggunaan robot dan chatbot untuk interaksi pelanggan, misalnya dalam perbankan atau industri perhotelan.

Sistem Sensor dan Robotika memiliki peran penting dalam otomatisasi, pengumpulan data, meningkatkan efisiensi, dan meningkatkan pengalaman pengguna dalam berbagai bidang.

- 5) **Komputer Visi (*Computer Vision*)** merupakan bidang dalam kecerdasan buatan yang bertujuan untuk memberikan kemampuan kepada komputer untuk memahami dan menganalisis gambar dan video. Tujuan utama dari *computer vision* adalah untuk menginterpretasikan data visual dan memungkinkan komputer untuk "melihat" dunia seperti manusia.

Proses *computer vision* melibatkan sejumlah tahap pemrosesan yang kompleks, termasuk pemrosesan gambar, deteksi objek, ekstraksi fitur, pengenalan pola, dan pengambilan keputusan. Teknik-teknik dalam *computer vision* mencakup segmentasi gambar, deteksi objek, pengenalan wajah, pelacakan objek, rekonstruksi 3D, dan lain sebagainya.

Manfaat dari computer vision adalah sebagai berikut:

- a) Pengenalan Objek dan Deteksi

*Computer vision* dapat digunakan untuk mendeteksi dan mengenali objek dalam gambar dan video. Hal ini berguna dalam aplikasi keamanan, pengawasan lalu lintas, analisis medis, dan lain-lain.

- b) Pelacakan Objek

*Computer vision*, dapat digunakan untuk pelacakan objek bergerak dalam video. Hal ini bermanfaat dalam pemantauan, pengawasan, dan analisis gerak objek, seperti pelacakan kendaraan atau deteksi kecelakaan lalu lintas.

- c) Pengenalan Wajah

*Computer vision* dapat digunakan untuk mengenali wajah manusia dalam gambar dan video. Hal ini diterapkan dalam berbagai aplikasi, seperti keamanan, verifikasi identitas, pengenalan emosi, dan pengenalan wajah dalam foto grup.

- d) *Augmented Reality (AR)* dan *Virtual Reality (VR)*

*Computer vision* digunakan dalam pengembangan aplikasi AR dan VR, yang menggabungkan dunia nyata dan dunia digital. Dengan menggunakan *computer vision*, objek nyata dapat diidentifikasi dan interaksi antara objek nyata dan virtual dapat terjadi.



e) Pengolahan Citra Medis

*Computer vision* berperan penting dalam analisis dan diagnosis medis melalui pemrosesan gambar medis, seperti pemindaian MRI, CT scan, atau pemindaian ultrasound. Hal ini membantu dalam identifikasi penyakit, deteksi tumor, analisis jaringan, dan lain-lain.

f) Kendali Otomatis dan Robotika

*Computer vision* digunakan dalam sistem kendali otomatis dan robotika untuk mendeteksi dan memperoleh informasi tentang lingkungan sekitar. Hal ini memungkinkan robot untuk berinteraksi dengan objek dan lingkungan secara efisien.

#### **2.4. Mesin Pembelajaran (*Machine Learning*)**

Tahun 1959, Arthur Samuel pertama kali menggunakan istilah machine learning. Arthur Samuel mengatakan bahwa machine learning adalah bidang ilmu komputer yang memberikan komputer kemampuan pembelajaran tanpa pemrograman yang jelas. *Machine Learning* dapat diartikan sebagai metode komputasi yang didasarkan pada pengalaman untuk meningkatkan performa atau membuat prediksi yang akurat (Mohri *et al.* 2018). *Machine Learning* bertujuan untuk memahami prinsip-prinsip dasar belajar sebagai proses komputasi.

Salah satu komponen kecerdasan buatan adalah pembelajaran mesin, yang memungkinkan komputer untuk mempelajari dan melakukan pekerjaan secara otomatis (Reddy *et al.* 2018). Algoritma tertentu digunakan untuk melakukan proses pembelajaran mesin, yang memungkinkan komputer untuk secara otomatis melakukan tugas yang diberikan kepadanya. *Machine learning* dapat diterapkan ke berbagai bidang komputasi untuk merancang dan memprogram secara eksplisit algoritma dengan keluaran kinerja tinggi, misalnya, pemfilteran email spam, deteksi penipuan di jejaring media sosial, perdagangan saham *online*, deteksi wajah dan bentuk, diagnosis bidang kesehatan/medis, prediksi lalu lintas, karakter *recognition* dan rekomendasi produk lain (Alzubi *et al.* 2018). Pembelajaran mesin biasanya digunakan untuk mengklasifikasikan satu masalah menjadi beberapa

kelompok. Orang dapat dengan mudah mengidentifikasi objek dalam kehidupan sehari-hari, tetapi tidak selalu dapat menjelaskan dengan jelas. Karena itu, machine learning diperlukan untuk mengenali, mengidentifikasi, dan memprediksi data.

*Machine learning* dikategorikan menjadi sepuluh jenis berdasarkan algoritma melakukan *training* dan berdasarkan ketersediaan *output* selama *training*, antara lain: *supervised learning*, *semi-supervised learning*, *unsupervised learning*, *reinforcement learning*, *transduction* dan *learning to learn* (Khan *et al.* 2010; Al-jarrah *et al.* 2015; Dey 2016; Sandhya dan Raina 2016; Society 2017).

### 1) *Supervised Learning*

*Supervised learning* merupakan salah satu pendekatan dalam ML di mana model atau algoritma komputer belajar dari contoh-contoh data yang diketahui labelnya. Setiap contoh data dalam *supervised learning* digunakan untuk melatih model memiliki label atau target yang sudah diketahui, dan model berusaha untuk mempelajari pola atau hubungan antara fitur-fitur (*features*) dalam data dengan label yang sesuai. Algoritma ini membangkitkan suatu fungsi yang memetakan input ke *output* yang diinginkan. Kualitas hasil pembelajaran sangat tergantung pada kesesuaian *input* dan *output* yang diberikan. Dengan demikian, *user* sangat berperan dalam memvalidasi *input* dan *output* tersebut. Algoritma *supervised learning* biasanya digunakan untuk menyelesaikan masalah klasifikasi maupun regresi.

Manfaat dari *supervised learning* meliputi kemampuan untuk melakukan klasifikasi, regresi, deteksi anomali, dan lain sebagainya. Penerapan *supervised learning* meliputi berbagai bidang, seperti pengenalan wajah, klasifikasi dokumen, analisis sentimen, prediksi harga saham, dan diagnosa medis. Model *supervised learning*, dapat belajar dari data yang ada untuk menghasilkan prediksi yang akurat dan berdasarkan pola yang teridentifikasi. Tahapan yang dilakukan dalam *supervised learning* antara lain:

#### a) Pengumpulan Data

Data yang dikumpulkan harus memiliki label atau target yang diketahui. Misalnya, untuk mengembangkan model klasifikasi email menjadi "spam"

atau "non-spam", diperlukan kumpulan data yang terdiri dari email beserta label yang menandakan apakah email tersebut adalah spam atau bukan.

b) Pembagian Data Latih dan Data Uji

Data yang telah dikumpulkan akan dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*test data*). Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk menguji kinerja model yang telah dilatih.

c) Pemilihan Model

Berbagai jenis model atau algoritma dapat digunakan dalam *supervised learning*, seperti *decision trees*, *support vector machines*, *neural networks*, dan lain-lain. Pemilihan model bergantung pada karakteristik data dan tujuan dari masalah yang ingin diselesaikan.

d) Pelatihan Model

Pada tahap pelatihan, model diberikan data latih dan melakukan iterasi untuk menyesuaikan parameter atau bobotnya. Model mencoba mempelajari pola dalam data latih dan membangun hubungan antara fitur-fitur dengan label yang sesuai.

e) Evaluasi Model

Setelah model dilatih, kinerja model akan dievaluasi menggunakan data uji yang terpisah. Metrik evaluasi yang digunakan antara lain akurasi, presisi, *recall*, dan *F1-score*. Evaluasi ini memberikan gambaran tentang seberapa baik model dapat memprediksi label dengan benar.

f) Penggunaan Model

Setelah model dievaluasi dan memperoleh kinerja yang memadai, model dapat digunakan untuk melakukan prediksi pada data baru yang tidak diketahui. Model akan menerima fitur-fitur dari data baru dan menghasilkan prediksi berdasarkan apa yang telah dipelajari selama pelatihan.

2) *Unsupervised Learning*

*Unsupervised learning* merupakan salah satu pendekatan ML yang modelnya belajar dari data tanpa adanya label atau target yang diketahui sebelumnya.

Dalam *unsupervised learning*, tujuan utamanya adalah menemukan pola, struktur, atau hubungan yang tersembunyi dalam data tanpa diketahui apa yang harus dicari. Algoritma ini memodelkan sekumpulan *input* secara otomatis tanpa ada panduan (yang berupa *output* yang diinginkan). Hal ini berarti bahwa data-data yang dipelajari hanya berupa *input* tanpa label kelas. Algoritma *unsupervised learning* biasanya digunakan untuk mengatasi masalah klasterisasi (*clustering*).

Manfaat dari *unsupervised learning* meliputi kemampuan untuk menemukan pola yang tidak diketahui sebelumnya dalam data, menggambarkan data dengan cara yang lebih ringkas, dan mengidentifikasi asosiasi atau hubungan yang tersembunyi. Penerapannya meliputi segmentasi pasar, pengelompokan konsumen, pengenalan pola dalam citra atau teks, dan deteksi anomali dalam data keamanan. Penerapan *unsupervised learning*, dapat diperoleh wawasan baru dan pengetahuan yang berharga dari data tanpa adanya label atau target yang diketahui. Karakteristik dari model *unsupervised learning*, antara lain:

a) Data Tanpa Label

Data yang digunakan dalam *unsupervised learning* tidak memiliki label atau target yang diketahui sebelumnya. Data ini terdiri dari fitur-fitur atau atribut yang dapat digunakan untuk menggambarkan karakteristik atau sifat data tersebut.

b) Pemrosesan Data

Tahap awal dalam *unsupervised learning* adalah pemrosesan data untuk memastikan kualitas dan kecocokan data. Tahapan ini melibatkan penghapusan data yang hilang atau tidak relevan, normalisasi data, atau pemrosesan lainnya untuk memastikan data siap digunakan.

c) *Clustering* (Pengelompokan)

Salah satu tugas utama dalam *unsupervised learning* adalah *clustering* atau pengelompokan. Tujuan *clustering* adalah mengelompokkan data ke dalam kelompok atau *cluster* berdasarkan kesamaan fitur atau sifat tertentu. Algoritma *clustering* seperti K-means, *Hierarchical Clustering*, atau DBSCAN digunakan untuk melakukan pengelompokan.

d) Pengurangan Dimensi

*Unsupervised learning* melibatkan pengurangan dimensi, dari data yang kompleks untuk menggambarkan data dengan cara yang lebih ringkas. Pengurangan dimensi dapat membantu mengidentifikasi fitur-fitur penting atau menghilangkan noise dalam data. Metode pengurangan dimensi yang umum digunakan termasuk *Principal Component Analysis (PCA)* dan *t-SNE (t-Distributed Stochastic Neighbor Embedding)*.

e) Asosiasi dan Anomali

*Unsupervised learning* digunakan untuk menemukan asosiasi atau hubungan antara item dalam data. Contohnya adalah *apriori algorithm* yang digunakan untuk menemukan aturan asosiasi dalam data transaksi. Selain itu, algoritma *unsupervised learning* dapat digunakan untuk mendeteksi anomali atau data yang berbeda atau tidak biasa dari yang diharapkan.

f) Visualisasi Data

*Unsupervised learning* dapat membantu dalam visualisasi data dengan cara yang bermakna. Penggunaan algoritma seperti algoritma reduksi dimensi dan teknik visualisasi seperti *scatter plot* atau *network visualization*, data kompleks dapat digambarkan secara visual untuk memahami pola dan struktur yang ada.

3) *Semi-Supervised Learning*

*Semi-Supervised Learning* merupakan paradigma ML yang sebagian besar data pelatihan tidak memiliki label, sedangkan sebagian kecil datanya memiliki label. Algoritma ini mengkombinasikan kedua algoritma *supervised learning* dan *unsupervised learning*, yang berisikan data sampel-sampel *input* yang berlabel dan tidak berlabel. Algoritma ini membangkitkan suatu fungsi atau pengklasifikasi yang tepat berdasarkan semua sampel *input* yang diberikan.

Perbedaan utama antara *semi-supervised learning*, *supervised learning*, dan *unsupervised learning* adalah dalam hal ketersediaan label pada data pelatihan. Dalam *supervised learning*, model ML mempelajari pola berdasarkan

pasangan data-input dan label yang diketahui. Model ini kemudian dapat menggunakan pengetahuannya untuk memprediksi label baru untuk data yang belum pernah dilihat sebelumnya. Di sisi lain, dalam pembelajaran tanpa pengawasan, model harus menemukan pola dan struktur di dalam data tanpa adanya label yang diketahui sebelumnya. Hal ini dapat dilakukan dengan teknik seperti *clustering* dan reduksi dimensi. *Semi-supervised learning* memanfaatkan data yang memiliki label untuk membantu dalam mempelajari pola pada data yang tidak berlabel.

#### 4) *Reinforcement Learning*

*Reinforcement learning* merupakan cabang pembelajaran mesin melalui proses belajar agen untuk berinteraksi dengan lingkungannya dan mengambil tindakan untuk mencapai tujuan tertentu. Proses yang dilakukan pada metode ini, agen tidak diberi contoh data pelatihan atau label, tetapi secara iteratif belajar melalui percobaan dan pengalaman. Algoritma ini mempelajari suatu kebijakan untuk melakukan aksi berdasarkan hasil pengamatan terhadap lingkungan yang ada. Setiap aksi menghasilkan akibat bagi lingkungan tersebut, dan lingkungan memberikan umpan balik (*feedback*) untuk memandu algoritma tersebut.

*Reinforcement learning* melibatkan tiga elemen utama antara lain agen, lingkungan, dan kebijakan. Agen adalah entitas yang belajar dan mengambil tindakan dalam lingkungan. Lingkungan adalah tempat agen beroperasi, dan dapat berupa simulasi komputer, permainan, robot fisik, atau bahkan dunia nyata. Kebijakan (*policy*) adalah strategi yang digunakan oleh agen untuk memilih tindakan berdasarkan keadaan lingkungan yang diberikan.

Proses pembelajaran dalam *reinforcement learning* melibatkan serangkaian langkah-langkah sebagai berikut:

##### a) Persepsi

Agen menerima informasi tentang keadaan (*state*) dari lingkungan. Keadaan dapat mencakup data sensorik seperti gambar, nilai-nilai sensor,

atau atribut lain yang menggambarkan keadaan lingkungan pada saat tertentu.

b) Tindakan

Agen memilih tindakan yang akan diambil berdasarkan kebijakan yang ada. Tindakan tersebut mempengaruhi keadaan lingkungan.

c) Umpan balik

Agen menerima umpan balik dari lingkungan dalam bentuk hadiah atau hukuman (*reward* atau *punishment*) setelah mengambil tindakan tertentu. Hadiah ini memberikan sinyal kepada agen tentang apakah tindakan yang diambil benar atau salah dalam mencapai tujuan yang diinginkan.

d) Pembelajaran

Agen memperbarui kebijakan berdasarkan umpan balik yang diterima. Tujuan agen adalah untuk memaksimalkan total hadiah yang diterima seiring berjalannya waktu.

Penerapan *reinforcement learning* sangat luas dan telah digunakan dalam berbagai bidang, antara lain:

a) Permainan

*Reinforcement learning* telah digunakan untuk melatih agen komputer dalam permainan strategi seperti catur, Go, dan poker. AlphaGo yang dikembangkan oleh DeepMind adalah contoh sukses dari penerapan *reinforcement learning* dalam permainan.

b) Robotika

*Reinforcement learning* dapat digunakan untuk melatih robot untuk melakukan tugas-tugas seperti berjalan, menghindari rintangan, atau merencanakan pergerakan yang kompleks.

c) Kendali Sistem

*Reinforcement learning* digunakan untuk mengoptimalkan sistem kendali yang kompleks, seperti pengendalian lalu lintas, pengelolaan sumber daya, atau pengaturan jaringan komunikasi.

d) Keuangan

*Reinforcement learning* digunakan dalam pengambilan keputusan investasi, manajemen risiko, dan pengoptimalan portofolio.

e) Penyempurnaan Otomatis

*Reinforcement learning* digunakan untuk melakukan penyempurnaan otomatis pada sistem, dalam lingkungan yang kompleks dan bervariasi seperti penyempurnaan otomatis dalam industri manufaktur.

Penerapan *reinforcement learning* terus berkembang dan menunjukkan potensi besar dalam banyak bidang di mana agen perlu belajar dan beradaptasi dengan lingkungannya untuk mencapai tujuan yang diinginkan.

5) *Transduction*

*Transduction* merupakan konsep dalam *machine learning* yang mengacu pada proses memprediksi label untuk contoh-contoh yang spesifik, biasanya dalam konteks *semi-supervised learning*. Tujuan transduksi adalah memanfaatkan informasi dari data berlabel dan data tanpa label untuk memperoleh prediksi yang akurat untuk contoh-contoh yang belum diberi label. Algoritma ini mirip dengan *supervised learning*, tetapi tidak secara eksplisit membangun suatu fungsi. Algoritma ini digunakan untuk berlatih memprediksi *output* baru berdasarkan *training inputs*, *training outputs*, dan *testing inputs* yang tersedia selama proses pembelajaran (pelatihan).

Perbedaan utama antara transduksi dan induksi (biasa dikenal sebagai prediksi) adalah bahwa transduksi berfokus pada prediksi yang akurat, sedangkan induksi bertujuan untuk membangun model yang dapat melakukan prediksi yang akurat. Penerapan transduksi dalam *machine learning* terus berkembang, dan teknik yang digunakan dapat bervariasi tergantung pada konteks dan jenis data yang dihadapi. Tujuan utamanya adalah untuk memanfaatkan informasi yang ada untuk memberikan prediksi yang akurat pada contoh-contoh yang belum diberi label. Penerapan transduksi dapat ditemukan dalam berbagai bidang dalam *machine learning*, antara lain:



a) *Semi-Supervised Learning*

Penerapan *semi-supervised learning* dapat memanfaatkan informasi dari contoh-contoh berlabel untuk memperbaiki kualitas model dan memberikan prediksi yang lebih baik untuk contoh-contoh tanpa label.

b) Pemrosesan Bahasa Alami

Transduksi dapat digunakan untuk memprediksi tag kata, entitas bernama, atau fitur lainnya pada teks yang belum diberi label. Pemrosesan bahasa alami memanfaatkan informasi dari teks yang telah diberi label, model dapat melakukan transduksi untuk memberikan label yang tepat pada teks yang tidak berlabel.

c) Deteksi Anomali

Transduksi dalam deteksi anomali dapat digunakan untuk memprediksi apakah contoh yang diberikan termasuk dalam kategori normal atau anomali. Dengan memanfaatkan contoh-contoh yang diberi label sebagai panduan, dapat dilakukan transduksi untuk mengidentifikasi contoh-contoh anomali yang belum diberi label.

d) Analisis Citra dan Pengenalan Objek

Analisis citra dan pengenalan objek, transduksi dapat digunakan untuk memberikan prediksi label pada objek dalam citra yang belum diberi label. Dengan menggunakan teknik transduksi, model dapat memanfaatkan contoh-contoh yang diberi label untuk memperoleh prediksi yang lebih baik untuk objek yang tidak berlabel dalam citra.

6) *Learning to Learn*

*Learning to Learn* merupakan paradigma dalam ML yang bertujuan untuk mengembangkan model atau algoritma yang dapat secara otomatis mempelajari strategi dan penyesuaian untuk memperbaiki kinerja pembelajarannya sendiri. *Learning to learn*, model tidak hanya belajar dari data, tetapi juga belajar cara belajar yang lebih efisien dan efektif. Algoritma ini mempelajari bias induktifnya sendiri berdasarkan pengalaman sebelumnya.

Konsep dasar di balik *learning to learn* adalah bahwa model atau agen pembelajaran mesin harus mampu menggeneralisasi dari pengalaman pembelajaran sebelumnya untuk menghadapi tugas baru dan mempelajari dengan cepat. Tujuannya adalah meminimalkan ketergantungan pada data pelatihan yang besar dan meningkatkan adaptasi terhadap perubahan lingkungan atau tugas. Beberapa pendekatan *Learning to Learn*, adalah sebagai berikut:

a) *Meta-Learning*

Pendekatan *meta-learning* melibatkan pembelajaran strategi atau algoritma yang dapat digunakan oleh model untuk mengatur proses pembelajaran. Model mempelajari cara memilih atau menyesuaikan arsitektur model, hiperparameter, atau algoritma pembelajaran yang paling sesuai untuk tugas tertentu berdasarkan pengalaman pembelajaran sebelumnya. Contohnya adalah pembelajaran otomatis hiperparameter (*automated hyperparameter tuning*) dan optimasi arsitektur *neural network* (*neural architecture search*).

b) Pembelajaran *Hierarchy*

Pendekatan ini mencoba mempelajari representasi hierarkis yang memungkinkan model untuk memperoleh pemahaman yang lebih abstrak dan umum dari data. Dalam hierarki ini, model dapat mempelajari penyesuaian atau transfer dari tugas-tugas yang lebih rendah ke tugas yang lebih tinggi tingkatannya. Contohnya adalah *Hierarchical Reinforcement Learning* yang menggunakan hierarki kebijakan (*policy*) untuk mempelajari tindakan pada berbagai tingkatan abstraksi.

c) Pembelajaran Penyesuaian Cepat (*Fast Adaptation*)

Pendekatan ini mengacu pada pembelajaran model yang dapat dengan cepat menyesuaikan diri dengan tugas baru berdasarkan pengalaman sebelumnya. Model memperbarui parameter atau pengetahuannya selama fase adaptasi untuk mengoptimalkan kinerja pada tugas yang baru. Contohnya adalah metode pembelajaran metrik (*metric learning*) yang memperbarui jarak atau pengukuran kesamaan antara contoh-contoh data.

Penerapan *learning to learn* sangat luas dan terus berkembang dalam berbagai aspek *machine learning*. Tujuan utamanya adalah untuk meningkatkan kemampuan model untuk belajar secara efisien, beradaptasi dengan cepat, dan mengoptimalkan performa pada tugas baru atau dalam kondisi yang berubah. Beberapa contoh penerapan *learning to learn* antara lain:

a) Optimasi Hiperparameter

Dalam *machine learning*, banyak algoritma dan model memiliki hiperparameter yang harus disesuaikan secara manual. Dengan menggunakan *learning to learn*, dapat mempelajari strategi untuk mengatur hiperparameter dengan efisien. Misalnya, algoritma pembelajaran meta dapat diterapkan untuk melatih model dalam memilih nilai-nilai yang optimal untuk hiperparameter seperti laju pembelajaran, jumlah iterasi, atau ukuran *batch*.

b) *Neural Architecture Search*

Membangun arsitektur *neural network* yang optimal adalah tugas yang membutuhkan banyak percobaan dan penyesuaian. Proses *learning to learn*, dapat menggunakan metode seperti *Neural Architecture Search* (NAS) untuk mempelajari strategi yang efisien dalam mencari arsitektur *neural network* yang baik. Model dapat mempelajari cara membangun dan menyesuaikan struktur jaringan berdasarkan performa dan pengalaman sebelumnya.

c) *Transfer Learning*

*Learning to learn* dapat diterapkan dalam *transfer learning* untuk memperbaiki adaptasi dari pengetahuan yang telah dikuasai pada tugas-tugas baru. Dengan mempelajari strategi adaptasi yang efisien, model dapat mentransfer pengetahuan secara lebih baik dan meningkatkan kinerja pada tugas baru dengan data yang terbatas.

d) Pembelajaran *Meta-Reinforcement*

*Reinforcement Learning*, *learning to learn* dapat digunakan untuk melatih agen untuk belajar secara cepat dan efisien dalam berbagai tugas. Model dapat mempelajari kebijakan yang dapat menyesuaikan diri dengan cepat

berdasarkan pengalaman sebelumnya, memungkinkan agen untuk belajar lebih efisien dalam menjelajahi lingkungan dan mencapai tujuan.

e) Pembelajaran *Active Learning*

*Learning to learn* dapat digunakan dalam pembelajaran aktif (*active learning*), di mana model belajar untuk memilih sampel data yang paling informatif untuk dilabeli oleh pengawas. Dengan mempelajari strategi pemilihan sampel yang efisien, model dapat mempercepat proses pembelajaran dengan mengurangi jumlah sampel yang harus dilabeli.

## 2.5. Kecerdasan Buatan untuk Prediksi *Narcissistic Personality Disorder*

Penerapan kecerdasan buatan untuk prediksi *narcissistic personality disorder* adalah area yang sedang berkembang dalam penelitian. *Narcissistic personality disorder* adalah gangguan mental yang ditandai oleh perilaku dan pola pikir yang berlebihan mengenai kepentingan diri sendiri, kebutuhan akan pujian dan pengakuan, serta kurangnya empati terhadap orang lain. Prediksi *narcissistic personality disorder* dapat membantu dalam mengidentifikasi individu yang berisiko mengalami gangguan ini, sehingga memungkinkan intervensi dan pengobatan dini. Penerapan kecerdasan buatan untuk prediksi *narcissistic personality disorder* dapat melibatkan berbagai metode dan teknik, antara lain:

a) Analisis Teks

Kecerdasan buatan dapat digunakan untuk menganalisis teks yang dihasilkan oleh individu, seperti tulisan di media sosial, blog, atau catatan medis. Metode pemrosesan bahasa alami (*Natural Language Processing: NLP*) dapat digunakan untuk mengidentifikasi pola linguistik atau penggunaan kata-kata yang konsisten dengan ciri-ciri *narcissistic personality disorder*.

b) Analisis Suara

Kecerdasan buatan dapat menganalisis karakteristik suara individu melalui teknik pengolahan suara. Beberapa penelitian telah mencoba mengidentifikasi pola suara yang berkaitan dengan ciri-ciri *narcissistic personality disorder*, seperti intonasi yang dominan atau kecenderungan berbicara tentang diri sendiri dengan cara yang berlebihan.

c) Analisis Gambar dan Video

Kecerdasan buatan juga dapat diterapkan dalam menganalisis gambar atau video individu untuk mencari tanda-tanda visual yang berkaitan dengan *narcissistic personality disorder*. Misalnya, ekspresi wajah yang terlalu percaya diri atau bahasa tubuh yang dominan dapat menjadi petunjuk indikator potensial *narcissistic personality disorder*.

d) Data Multimodal

Pendekatan yang lebih maju melibatkan penggabungan informasi dari berbagai sumber data, seperti teks, suara, dan gambar. Metode kecerdasan buatan yang lebih kompleks seperti *deep learning* dapat digunakan untuk memodelkan ketergantungan dan hubungan antara data multimodal ini dalam konteks prediksi *narcissistic personality disorder*.

Prediksi *narcissistic personality disorder* menggunakan kecerdasan buatan melibatkan beberapa tahap yang mencakup pengumpulan data, pemrosesan data, pembangunan model, dan evaluasi model. Berikut adalah proses prediksi *narcissistic personality disorder* menggunakan kecerdasan buatan:

a) Pengumpulan Data

Langkah pertama dalam memprediksi *narcissistic personality disorder* adalah mengumpulkan data yang relevan. Data ini dapat mencakup berbagai sumber, seperti teks dari media sosial, catatan medis, hasil tes psikologis, atau wawancara. Data ini harus mencakup informasi tentang perilaku, pola pikir, dan karakteristik individu yang dapat digunakan untuk mengidentifikasi tanda-tanda *narcissistic personality disorder*.

b) Pemrosesan Data

Setelah data terkumpul, langkah berikutnya adalah memproses data tersebut. Pemrosesan data melibatkan langkah-langkah seperti pembersihan data, normalisasi, dan ekstraksi fitur. Data dapat diubah menjadi representasi numerik yang dapat dimengerti oleh algoritma kecerdasan buatan. Selain itu, fitur yang relevan dengan *narcissistic personality disorder* dapat diekstraksi, seperti penggunaan kata-kata yang berlebihan, kecenderungan untuk memuji diri sendiri, atau kurangnya empati.

c) Pembangunan Model kecerdasan buatan

Setelah data diproses, langkah selanjutnya adalah membangun model kecerdasan buatan untuk memprediksi *narcissistic personality disorder*. Berbagai metode kecerdasan buatan dapat digunakan, seperti *machine learning* dan *deep learning*. Beberapa algoritma yang umum digunakan dalam penerapan ini termasuk *decision tree*, *logistic regression*, *random forest*, atau *neural network*. Model ini akan belajar dari data yang diberikan dan mencari pola atau hubungan yang dapat digunakan untuk memprediksi *narcissistic personality disorder*.

d) Pelatihan Model

Setelah model kecerdasan buatan dibangun, tahap berikutnya adalah melatih model menggunakan data yang telah diproses. Data ini akan dibagi menjadi subset pelatihan dan pengujian. Model akan belajar dari subset pelatihan dan mengoptimalkan parameter internalnya untuk meningkatkan kinerja prediksi. Pelatihan model ini melibatkan iterasi berulang hingga model mencapai kinerja yang diinginkan.

e) Validasi dan Evaluasi Model

Setelah model dilatih, tahap selanjutnya adalah melakukan validasi dan evaluasi untuk mengukur kinerja model. Model akan diuji menggunakan subset pengujian yang tidak digunakan dalam pelatihan. Metrik evaluasi seperti akurasi, presisi, kepekaan, atau kekhususan digunakan untuk mengukur performa model. Validasi ini membantu menentukan sejauh mana model dapat memprediksi *narcissistic personality disorder* dengan akurasi dan keandalan yang memadai.

## 2.6. Database Google Scholar

Google scholar merupakan media yang menyediakan database publikasi ilmiah untuk pencarian jurnal-jurnal bereputasi secara online baik skala nasional dan internasional (Karim *et al.* 2021). Google scholar diluncurkan pada tahun 2004 oleh induk perusahaan Google (Setiani Rafika *et al.* 2017). Layanan Google Scholar menyediakan file berupa PDF (*Portable Document Format*) secara lengkap dan

gratis. *Database* Google Scholar menawarkan kemudahan dalam melakukan pencarian literatur akademis secara luas dan bebas. Peneliti dapat dengan mudah menemukan seluruh bidang ilmu dan referensi ilmiah dari *database* tersebut. Dokumen yang tersedia di Google Scholar meliputi makalah peer-reviewed, thesis, buku, abstrak, dan artikel dari penerbit akademis, komunitas profesional, pusat data pracetak, universitas, dan organisasi akademis lainnya.

Google Scholar bekerja dengan cara menemukan penelitian yang paling relevan dari seluruh penelitian akademis. Hasil paling relevan akan selalu muncul pada halaman pertama. Google Scholar memiliki peralatan canggih yang memungkinkan pelacak, analisis, dan visualisasi hasil penelitian dalam waktu hitungan detik tanpa koma. Tahun penelitian, penulis, kata-kata, penerbit, tahun terbit, dan kata-kata memungkinkan Google Scholar untuk membandingkan hasil penelitian.

## **2.7. Aplikasi Publish or Perish**

Pemanfaatan aplikasi *Publish or Perish* (PoP) merupakan salah satu cara yang digunakan untuk mendapatkan dan menganalisis referensi jurnal ilmiah akademik dalam penulisan karya. Tahun 2006, aplikasi *Publish or Perish* (PoP) mulai diperkenalkan kepada komunitas akademik dan publik umum. Aplikasi *Publish or Perish* (PoP) memanfaatkan *Google scholar query* untuk memperoleh informasi tentang jurnal ilmiah. *Google scholar query* akan menganalisis dan mengkonversi jurnal ilmiah kedalam beberapa perhitungan statistik. Aplikasi *Publish or Perish* (PoP) dapat berjalan di sistem operasi windows, machintosh dan linux yang terhubung terhadap jaringan internet. Aplikasi *Publish or Perish* (PoP) juga dapat menyalin hasil pencarian jurnal ilmiah yang tersedia ke dalam *clipboard windows*.

Aplikasi *Publish or Perish* (PoP) dikembangkan untuk membantu penulis atau peneliti akademisi untuk mencari dan memperoleh referensi jurnal ilmiah penelitian untuk penulisan karya ilmiah (Bilferi Hutapea 2023). Selain itu, aplikasi *Publish or Perish* (PoP) dapat digunakan untuk melakukan penelitian bibliometrik,

menentukan jurnal ilmiah mana yang akan digunakan sebagai referensi dan rujukan, dan melakukan tinjauan literatur. Penggunaan aplikasi Publish or Perish (PoP) terhadap jurnal ilmiah yang akan dijadikan referensi dapat menampilkan metrik seperti sitasi per penulis, sitasi rata-rata per artikel, sitasi per tahun, dan sitasi keseluruhan artikel yang terkait dengan pencarian. Dengan menggunakan aplikasi Publish or Perish (PoP), banyak tampilan metrik akan mempermudah penulis untuk memilih referensi jurnal ilmiah, seperti analisis jumlah penulis per artikel, h-index dan parameter yang terkait dengan g-index, peningkatan rata-rata tahunan h-index individu, tiga variasi h-index individu, dan tingkat kutipan tertimbang pada waktu tertentu. Aplikasi Publish or Perish (PoP) juga menawarkan akses metadata Crossref, Google Scholar, Google Scholar Profile, dan PubMed secara gratis. Selain itu, dengan melakukan registrasi untuk pencarian referensi jurnal ilmiah yang terintegrasi dengan OpenAlex, Scopus, Semantic Scholar, dan Web of Science, Aplikasi Publish or Perish (PoP) menawarkan akses metadata secara gratis.

## 2.8. Aplikasi VOSviewer

VOSviewer merupakan perangkat lunak untuk membantu peneliti dalam membuat peta penelitian berdasarkan *network data* dan untuk memvisualisasikan serta mengeksplorasi peta tersebut (Jan van Eck dan Waltman 2022). Vosviewer, juga dikenal sebagai VV, adalah sebuah perangkat lunak yang dapat digunakan untuk menampilkan peta bibliometrik atau set data yang mengandung field bibliografi seperti judul, pengarang, penulis, jurnal, dan lain-lain (Karim *et al.* 2021). VOSviewer dapat digunakan untuk membuat peta penulis atau jurnal berdasarkan data *co-citation* atau membuat peta kata kunci berdasarkan data *co-occurrence*. Aplikasi ini menawarkan penampil yang memungkinkan peta bibliometrik diperiksa secara lengkap. VOSviewer memiliki beberapa fungsi yang dapat dirangkum sebagai berikut:

1. Membangun peta berdasarkan *network data*. Peta dapat dibuat berdasarkan *network* yang telah tersedia, namun dapat pula dibangun jaringan terlebih dahulu. VOSviewer dapat digunakan untuk membangun *network* publikasi



ilmiah, jurnal ilmiah, peneliti, organisasi penelitian, negara, kata kunci, atau istilah. Item-item dalam *network* ini dapat dihubungkan melalui penulisan bersama, kejadian bersama, kutipan, penggabungan bibliografi, atau tautan kutipan bersama. Pembentukan *network* memerlukan *file database* bibliografi (yaitu file Web of Science, Scopus, Dimensions, Lens, dan PubMed) dan file manajer referensi (yaitu file RIS, EndNote, dan RefWorks) yang akan menjadi *inputan* dalam aplikasi VOSviewer.

2. Memvisualisasikan dan mengeksplorasi peta. Aplikasi VOSviewer mampu menampilkan hasil peta dalam tiga bentuk visualisasi peta yaitu *network visualization*, *overlay visualization*, dan *density visualization*. Aplikasi VOSviewer juga memiliki fasilitas zoom dan gulir yang berfungsi untuk menggali dan mengeksplorasi secara detail dan diperlukan saat peneliti menggunakan item yang berjumlah sangat banyak atau memiliki ribuan item.

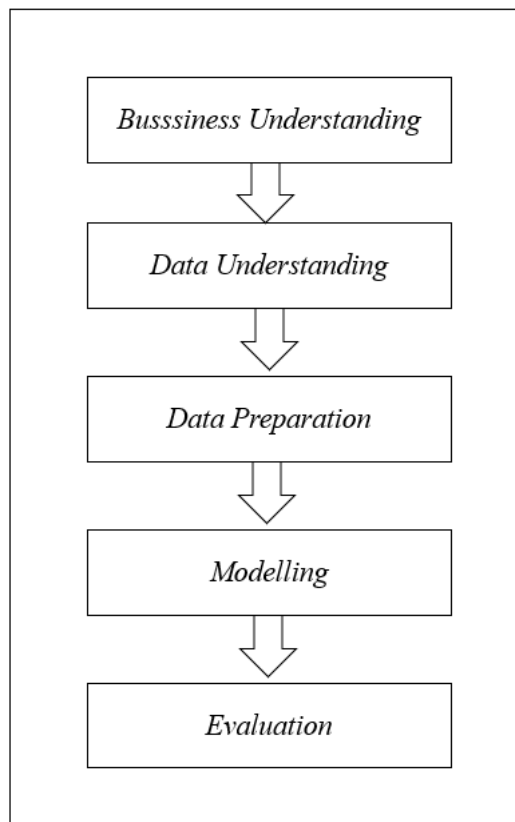
VOSviewer terutama dirancang untuk menganalisis jaringan bibliometrik, tetapi juga dapat membuat, memvisualisasikan, dan menjelajahi peta berdasarkan berbagai jenis data jaringan. Vosviewer dibuat dalam Java, sehingga berfungsi pada sebagian besar platform perangkat keras dan sistem operasi.

### **III. METODOLOGI PENELITIAN**

Untuk mencapai tujuan penelitian, tahapan penelitian harus disusun secara sistematis. Tahapan-tahapan penelitian tersebut menjadi acuan dalam melaksanakan penelitian. Pada bab ini dijelaskan mengenai tahapan-tahapan penelitian dan kerangka pikir dari penyiapan data hingga pengembangan model dan perbandingan model klasifikasi.

#### **3.1. Tahapan Penelitian**

Penelitian ini menggunakan tahapan model *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Tahapan-tahapan CRISP-DM, antara lain *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*. Namun dalam penelitian ini hanya sampai pada tahapan evaluasi. Adapun tahapan yang dilakukan dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

**1) Business Understanding**

*Business understanding* merupakan pemahaman terhadap tujuan bisnis, penilaian situasi dan menerjemahkan tujuan bisnis ke dalam tujuan data mining. Dalam penelitian ini membutuhkan pengetahuan untuk mendapatkan data penderita gangguan kepribadian terutama penderita *Narcissistic Personality Disorder* (NPD).

**2) Data Understanding**

Pada tahapan ini dilakukan proses pengumpulan data, kemudian melakukan analisa data serta melakukan evaluasi kualitas data yang digunakan dalam penelitian. Sumber data yang digunakan dalam penelitian ini merupakan data-data yang terdapat secara *online*. Metode yang dilakukan dalam rangka pengumpulan data yaitu studi pustaka. Metode pengumpulan data dengan studi kepustakaan ini dilakukan dengan mempelajari jurnal, dan buku-buku literatur yang berkaitan dengan *machine learning*, *narcissistic personality disorder*,

terutama literatur yang berhubungan dengan Data Mining, seleksi fitur, Algoritma *Random Forest Classifier*, Algoritma *Support Vector Machine* dan Algoritma Naive Bayes.

### 3) *Data Preparation*

Pada tahapan ini dilakukan pengumpulan data menggunakan data primer yang diperoleh secara *online* dari *Open Psychometrics* (<https://www.kaggle.com/datasets>). Dataset yang telah terkumpul disimpan dalam format .csv atau .arff agar dapat diproses lebih lanjut menggunakan Weka. Tampilan dataset yang digunakan dapat dilihat pada gambar 2.

1: gender	2: age	3: Q1	4: Q2	5: Q3	6: Q4	7: Q5	8: Q6	9: Q7	10: Q8	11: Q9	12: Q10	13: Q11	14: Q12	15: Q13	16: Q14	17: Q15	18: Q16	19: Q17
Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1.0	50.0	2.0	2.0	2.0	2.0	1.0	2.0	1.0	2.0	2.0	2.0	1.0	1.0	2.0	1.0	1.0	1.0	2.0
1.0	40.0	2.0	2.0	2.0	1.0	2.0	2.0	1.0	2.0	1.0	1.0	2.0	2.0	2.0	1.0	2.0	2.0	1.0
1.0	28.0	1.0	2.0	2.0	1.0	2.0	1.0	2.0	1.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0
1.0	37.0	1.0	1.0	2.0	2.0	2.0	1.0	2.0	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0
1.0	50.0	1.0	2.0	1.0	1.0	1.0	2.0	1.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0
1.0	27.0	1.0	2.0	2.0	1.0	2.0	1.0	1.0	1.0	2.0	2.0	1.0	1.0	1.0	2.0	1.0	1.0	1.0
1.0	45.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	1.0	2.0	1.0
2.0	36.0	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0
2.0	24.0	1.0	2.0	2.0	1.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	2.0	1.0	1.0
2.0	20.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0
1.0	25.0	2.0	2.0	2.0	1.0	2.0	1.0	1.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0
2.0	18.0	1.0	1.0	1.0	2.0	1.0	2.0	2.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	2.0	2.0	1.0
1.0	43.0	1.0	2.0	2.0	1.0	1.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	1.0	1.0
2.0	31.0	1.0	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	1.0	2.0
1.0	48.0	1.0	2.0	2.0	2.0	2.0	1.0	2.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0
1.0	36.0	1.0	2.0	2.0	1.0	2.0	2.0	1.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0
1.0	35.0	2.0	1.0	2.0	1.0	2.0	1.0	1.0	1.0	2.0	1.0	2.0	2.0	1.0	1.0	1.0	2.0	1.0
2.0	30.0	1.0	2.0	1.0	1.0	2.0	1.0	1.0	2.0	1.0	1.0	1.0	2.0	2.0	2.0	2.0	2.0	1.0
1.0	29.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0	1.0	2.0	2.0	1.0	1.0	2.0
1.0	37.0	2.0	2.0	2.0	1.0	1.0	2.0	1.0	1.0	2.0	1.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0
1.0	16.0	2.0	2.0	2.0	2.0	2.0	1.0	1.0	1.0	2.0	1.0	2.0	2.0	2.0	1.0	1.0	2.0	1.0
2.0	27.0	1.0	2.0	2.0	2.0	1.0	2.0	1.0	1.0	1.0	2.0	1.0	2.0	2.0	2.0	1.0	2.0	2.0
1.0	42.0	1.0	2.0	2.0	1.0	2.0	1.0	1.0	2.0	1.0	2.0	2.0	1.0	1.0	2.0	1.0	2.0	2.0
2.0	41.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0	1.0	2.0	2.0	1.0	1.0	2.0	2.0	1.0	1.0	2.0
1.0	37.0	1.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0	2.0	1.0	1.0	1.0
2.0	38.0	2.0	2.0	2.0	1.0	1.0	2.0	1.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	1.0	2.0	2.0

Gambar 2. Tampilan dataset *narcissistic personality disorder*

Dataset yang dikumpulkan memiliki jumlah *record* sebanyak 11.243, kemudian data diseleksi agar dapat digunakan dalam proses pemodelan. Data yang digunakan dalam penelitian ini memiliki ciri sebagai berikut:

- a) Jenis kelamin yang digunakan hanya male dan female, selain itu akan dihapus dari dataset (tidak digunakan dalam penelitian). Data awal jenis kelamin terdiri dari empat jenis yaitu male (1), female (2), other (3) dan tidak memilih (4).

- b) Rentang usia yang digunakan dalam penelitian ini yaitu usia 14 – 50 tahun. Pada data awal terdapat usia lebih dari 50 tahun, data tersebut tidak digunakan karena rentang usia yang paling sering terkena gangguan kepribadian yaitu rentang usia 14-50 tahun.
- c) Data jawaban responden terisi semua, karena pada data awal terdapat data pernyataan yang tidak diisi oleh responden (*missing value*).

Dari proses penyiapan data tersebut dan dari proses seleksi, terkumpul data sejumlah 8.340 data yang terdiri dari 44 fitur. Deskripsi fitur yang terdapat pada dataset dapat dilihat pada Tabel 2.

Tabel 2. Deskripsi Fitur

Fitur	Deskripsi
Jenis Kelamin ( <i>gender</i> )	Jenis kelamin yang digunakan yaitu male dan female
Usia ( <i>age</i> )	Rentang usia 14 – 50 tahun
Pernyataan 1 – 40 ( <i>question 1 – 40</i> )	Jawaban responden berupa nilai 1 atau 2 sesuai dengan kepribadian responden
Elapse	Merupakan selish waktu <i>submitted</i> dengan <i>loaded</i> dalam satuan detik
Skor ( <i>score</i> )	Penjumlahan hasil pengisian kuesioner responden
Kelas ( <i>class</i> )	Label kelas “ya” dan “tidak”

#### 4) *Modelling*

Dalam tahapan ini, berbagai macam teknik pemodelan dipilih dan diterapkan ke dataset yang sudah disiapkan untuk mengatasi permasalahan prediksi. Teknik yang digunakan yaitu teknik klasifikasi, seperti penerapan metode *Random Forest Classifier*, *Support Vector Machine* dan *Naive Bayes*. Selain itu, peneliti juga melakukan pemilihan fitur guna meningkatkan akurasi. Tujuan utama dari pemilihan fitur adalah untuk menghindari *overfitting* sehingga meningkatkan performa dari sistem dan sistem menghasilkan hasil yang lebih cepat dan efektif.

Teknik pemilihan fitur yang digunakan yaitu teknik *filter* seperti *information gain* dan *gain ratio*. Teknik pemilihan fitur tersebut berlandaskan pada sistem perangkaian menggunakan batasan nilai minimal informasi dari masing-masing fitur. Batasan nilai (*threshold*) yang digunakan pada penelitian ini yaitu  $>0,05$ . Hal ini berarti bahwa fitur yang memiliki nilai informasi lebih dari 0,05 akan digunakan untuk pengembangan model klasifikasi. Sedangkan fitur yang memiliki nilai informasi kurang dari 0,05 akan dibuang atau tidak digunakan dalam pengembangan model klasifikasi.

### 5) *Evaluation*

Pada tahapan ini, peneliti melakukan pengujian dan pengevaluasian nilai akurasi dari metode *machine learning*. Uji akurasi dilakukan untuk menentukan tingkat kepercayaan dari klasifikasi yang dihasilkan. Matriks konfusi digunakan untuk menguji akurasi interpretasi penelitian ini. Penerapan *confusion matrix* dapat mengidentifikasi dan menganalisis catatan dari berbagai kelas. Tabel *confusion matrix* ditunjukkan pada Tabel 3.

Tabel 3. *Confusion Matrix*

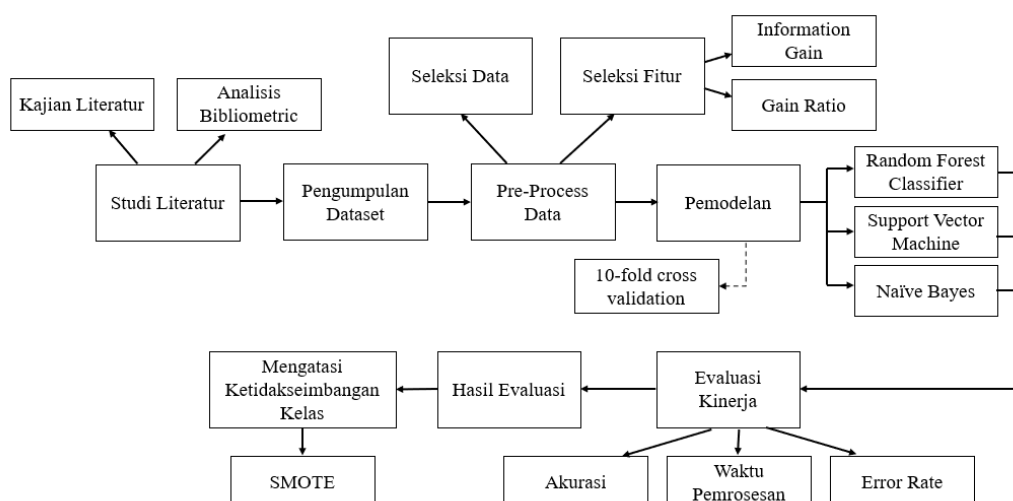
		Prediksi	
		Positif	Negatif
Aktual	Positif	TP	FN
	Negatif	FP	TN

Keterangan :

- a) TP (*True Positive*) merupakan banyaknya data yang kelas aktualnya adalah kelas positif dengan kelas prediksinya merupakan kelas positif.
- b) FN (*False Negative*) merupakan banyaknya data yang kelas aktualnya adalah kelas positif dengan kelas prediksinya merupakan kelas negatif.
- c) FP (*False Positive*) merupakan banyaknya data yang kelas aktualnya adalah kelas negatif dengan kelas prediksinya merupakan kelas positif.
- d) TN (*True Negative*) merupakan banyaknya data yang kelas aktualnya adalah kelas negatif dengan kelas prediksinya merupakan kelas negatif.

### 3.2. Diagram Alir Penelitian

Diagram alir penelitian adalah alir atau langkah yang menjadi landasan utama dalam sebuah penelitian. Dalam bab ini, akan dibahas tentang kerangka penelitian yang akan digunakan dalam penelitian ini. Diagram alir penelitian ini akan membantu dalam menguraikan kerangka kerja yang akan digunakan sebagai acuan pada seluruh proses penelitian ini. Diagram alir penelitian yang digunakan dalam penelitian ini dapat dilihat pada Gambar 3.



Gambar 3. Diagram alir penelitian

Gambar 3 menunjukkan kerangka berpikir yang dilakukan dalam penelitian ini. Diawali dengan melakukan studi literatur yang mengulas kajian literatur dan analisis bibliometric dari penelitian-penelitian terkait. Hasil analisis tersebut memetakan kata kunci dan topik saja yang sedang ramai dibahas. Setelah dilakukan studi literatur, dilanjutkan dengan pengumpulan data secara *online* yang diperoleh dari *Open Psychometrics* (<https://www.kaggle.com/datasets>). Dari dataset yang diperoleh, dilanjutkan pada tahapan pre-proses data yaitu melakukan seleksi data dan seleksi fitur agar data yang digunakan bebas noise dan mampu meningkatkan hasil akurasi. Pada proses pemilihan fitur, digunakan teknik yang banyak digunakan dalam penelitian yaitu teknik *information gain* dan *gain ratio*.

Setelah dilakukan pre-proses, tahapan selanjutnya yaitu melakukan pemodelan menggunakan tiga metode klasifikasi, yaitu metode *random forest classifier*, *support vector machine* dan *naive bayes*. Pada proses pemodelan, pembagian dataset menggunakan teknik *10-fold cross validation*. *10-fold cross-validation* adalah teknik penting dalam pengembangan dan evaluasi model *machine learning*. Selama proses *cross-validation*, model *machine learning* dilatih dan diuji sebanyak 10 kali, di mana setiap kali satu *fold* digunakan sebagai data uji dan sembilan *fold* lainnya digunakan sebagai data pelatihan.

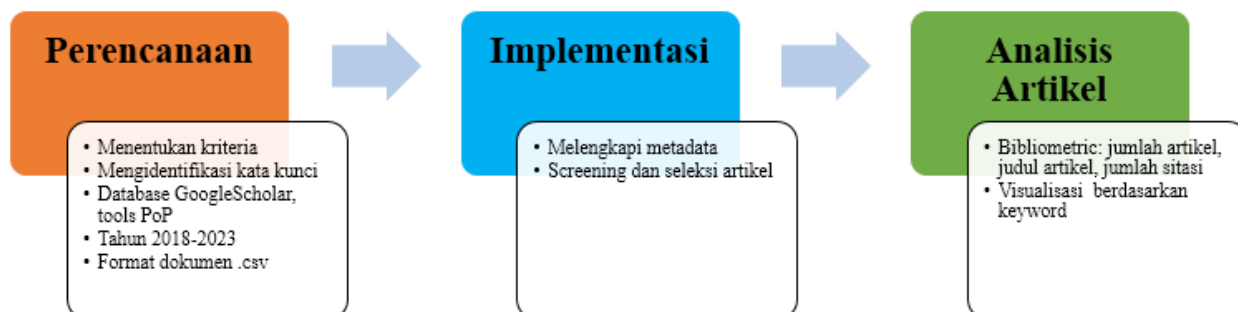
Tahapan selanjutnya yaitu melakukan evaluasi kinerja yang mengukur tingkat akurasi, lama waktu pemrosesan dan nilai *error rate*. Dari hasil evaluasi kinerja akan dilakukan perbandingan dari ketiga metode klasifikasi. Setelah dilakukan perbandingan akan diperoleh nilai akurasi tertinggi dan terendah. Hasil tersebut akan diproses selanjutnya untuk mengatasi ketidakseimbangan kelas yang terjadi pada dataset yang digunakan. Teknik yang digunakan untuk mengatasi ketidakseimbangan kelas yaitu SMOTE. Hal ini dilakukan agar mampu meningkatkan nilai akurasi dari metode pengklasifikasi.

### **3.2.1. Diagram Alir Penelitian Studi Literature Review**

Tahapan pada proses studi literature review ini memberikan kerangka kerja umum untuk melakukan *literature review* tentang kecerdasan buatan dalam diagnosa gangguan kepribadian. Pada penelitian ini dilakukan tinjauan literatur dengan menggunakan metode *systematic literature review (SLR)*. SLR bertujuan untuk mengumpulkan semua penelitian tentang topik tertentu, mengevaluasinya secara kritis, dan menyimpulkan hasil penelitian yang telah dianalisis (Budi dan Suryono 2023). SLR telah banyak digunakan dalam berbagai bidang penelitian antara lain bidang kesehatan (Whear *et al.* 2022; Alsaleh *et al.* 2023a; Ehiabhi dan Wang 2023; Jung *et al.* 2023; Wen *et al.* 2023), kecerdasan buatan (Alsaleh *et al.* 2023b; Saranya dan Subhashini 2023; de Wilde 2023), *machine learning* (Thieme *et al.* 2020; Bickel *et al.* 2023; Issah *et al.* 2023; Mintarya *et al.* 2023), pertanian (Sachithra dan



Subhashini 2023), keuangan/perbankan (Suryono *et al.* 2020) dan lain sebagainya. Tahapan yang dilakukan dalam membangun SLR dalam penelitian ini dilakukan dalam tiga tahapan yaitu perencanaan, implementasi dan pelaporan. Diagram alir proses SLR dapat dilihat pada Gambar 4.



Gambar 4. Diagram Alir Proses SLR

Secara detail, tahapan yang dilakukan adalah sebagai berikut:

#### 1) **Perencanaan**

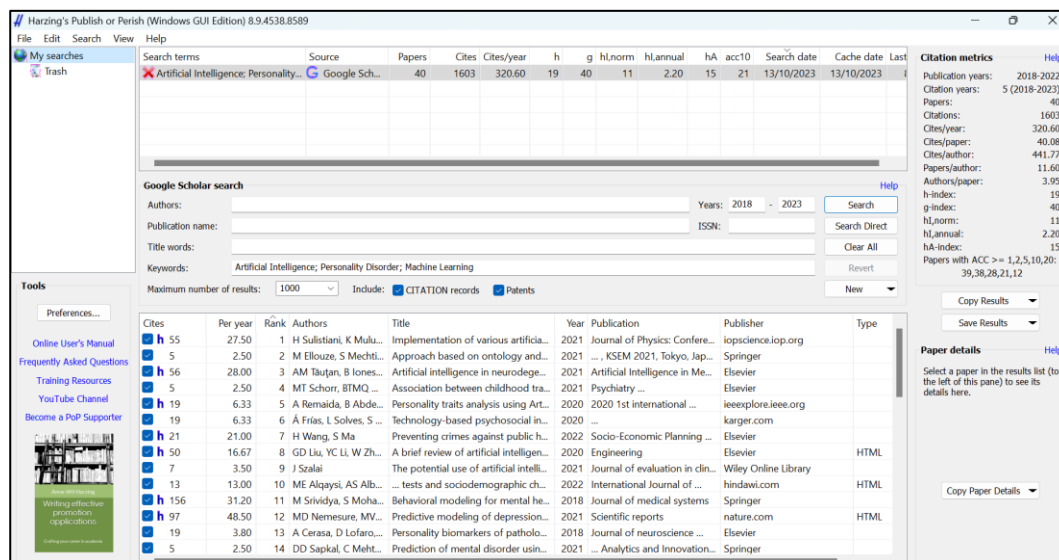
Langkah pertama yang dilakukan yaitu mengidentifikasi kebutuhan SLR seperti menentukan area topik penelitian, menentukan tujuan, menentukan kriteria pemilihan artikel dan mengidentifikasi kata kunci apa saja yang akan digunakan dalam mencari artikel penelitian. Sebelum melakukan SLR, seorang peneliti harus mengidentifikasi alasan melakukan SLR dan melihat apakah SLR merupakan strategi yang tepat. Selain itu, penting untuk melakukan pengecekan apakah hal yang sama pernah dilakukan di area tersebut sebelumnya. Jika sudah ada review sebelumnya, apakah SLR yang akan dilakukan memiliki justifikasi yang cukup kuat untuk tetap dilanjutkan. Selama proses pengerjaan SLR, peneliti harus membuat pertanyaan penelitian terlebih dahulu yang digunakan sebagai pegangan dalam pengerjaan SLR.

SLR penelitian ini bertujuan untuk menyelidiki berbagai metode kecerdasan buatan yang digunakan dalam diagnosa gangguan kepribadian. Selain itu, peneliti juga menentukan kriteria artikel pada proses *literature review* yang dapat dilihat pada Tabel 4.

Tabel 4. Kriteria artikel pada proses *literature review*

No	Kriteria Inklusi	Pengecualian
1	Artikel ditulis dalam bahasa Inggris dan diterbitkan dalam lima tahun terakhir (2018 – 2023)	Artikel tidak memuat pembahasan khusus tentang aplikasi dan metode kecerdasan buatan untuk gangguan kepribadian
2	Artikel membahas penerapan dan metode kecerdasan buatan dalam prediksi gangguan kepribadian	Artikel yang dipublikasikan merupakan pembahasan teori yang digunakan sebagai bahan kuliah atau tutorial
3	Artikel membahas berbagai jenis gangguan kepribadian dan cara penanganannya	Artikel hanya memuat abstrak atau bagian tertentu
4	Artikel yang terindeks <i>Scopus Quartile Q1-Q4</i>	

Pengumpulan data dilakukan melalui penelusuran publikasi yang terindeks *Google Scholar* menggunakan aplikasi *Publish or Perish (PoP)*. Kata kunci yang digunakan dalam penelitian ini adalah kata kunci yang relevan dengan topik SLR yang akan digunakan dalam pencarian literatur. Kata kunci yang digunakan antara lain “*Artificial Intelligence*”, “*Personality Disorder*”, “*Machine Learning*”. *Publish or Perish* dipilih karena aplikasi tersebut menyediakan berbagai macam pilihan *database* artikel, antara lain dari *Crossref*, *Google Scholar*, *PubMed*, *Microsoft Academic*, *Scopus*, dan *Web of Science*. Tampilan pengumpulan artikel menggunakan PoP dapat dilihat pada Gambar 5.



Gambar 5. Metode Penelusuran menggunakan PoP

*Database* yang digunakan dalam dalam proses pencarian artikel penelitian ini yaitu *Google Scholar*. Hal ini dikarenakan *Google Scholar* merupakan *database* yang dapat diakses secara bebas, terbuka untuk semua peneliti dan masyarakat umum, mencakup sebagian besar bidang penelitian ilmiah, dan merupakan sumber paling populer dari informasi ilmiah (Falagas *et al.* 2008). Pada proses awal pencarian berdasarkan kata kunci dan kriteria yang digunakan pada penelitian ini didapatkan sebanyak 1.022 artikel. Selanjutnya, data yang diperoleh akan disimpan dalam bentuk file RIS atau *Research Information Systems Citation File*.

## 2) Implementasi

Tahapan selanjutnya yaitu melengkapi metadata melalui Mendeley dan dilanjutkan dengan *screening* artikel. *Screening* dilakukan dengan cara membaca cepat pada bagian abstrak dan judul. Hasil dari proses *screening* menunjukkan bahwa terdapat 834 artikel yang tidak sesuai dengan kriteria pencarian penulis. Artikel-artikel tersebut tidak membahas tentang penerapan dan metode kecerdasan buatan dalam prediksi gangguan kepribadian. Artikel hanya melingkupi satu kata kunci saja, misal hanya membahas tentang *machine learning* saja atau hanya membahas tentang gangguan kepribadian saja.

Artikel yang tidak relevan atau tidak memenuhi kriteria penelitian akan dihapus dari list literatur. Hasil *screening* dapat dilihat pada Tabel 5.

Tabel 5. Hasil *screening* artikel

<b>Kriteria</b>	<b>Jumlah artikel</b>
Tidak relevan dengan kata kunci	834
Tidak dalam bahasa Inggris	14
Data rangkap	37
Tidak terindeks <i>Scopus</i>	115
Terindeks <i>Scopus quartile</i> Q3, Q4	8
Terindeks <i>Scopus quartile</i> Q1, Q2	14
<b>Total</b>	<b>1.022</b>

Total artikel yang diperoleh yaitu sebanyak 1.022 artikel, setelah melalui proses *screening* diperoleh 22 artikel yang terindeks *Scopus Quartile* Q1-Q4 yang akan digunakan dalam penelitian ini.

### 3) Analisis Artikel

Tahapan akhir dari SLR memasukkan data yang sudah diperoleh dari hasil *screening* ke aplikasi VOSviewer versi 1.6.19 dengan tujuan untuk memvisualisasikan pola jaringan atau hubungan antar bibliometrik ke dalam tiga kategori, diantaranya *network visualization*, *overlay visualization*, dan *density visualization*. *Network visualization* bertujuan untuk memvisualisasikan kuat atau tidaknya jaringan atau hubungan antar *term* (istilah) penelitian. *Overlay visualization* bertujuan untuk memvisualisasikan jejak historis berdasarkan tahun diterbitkannya penelitian, sedangkan *density visualization* bertujuan untuk menampilkan kerapatan atau penekanan pada kelompok penelitian.

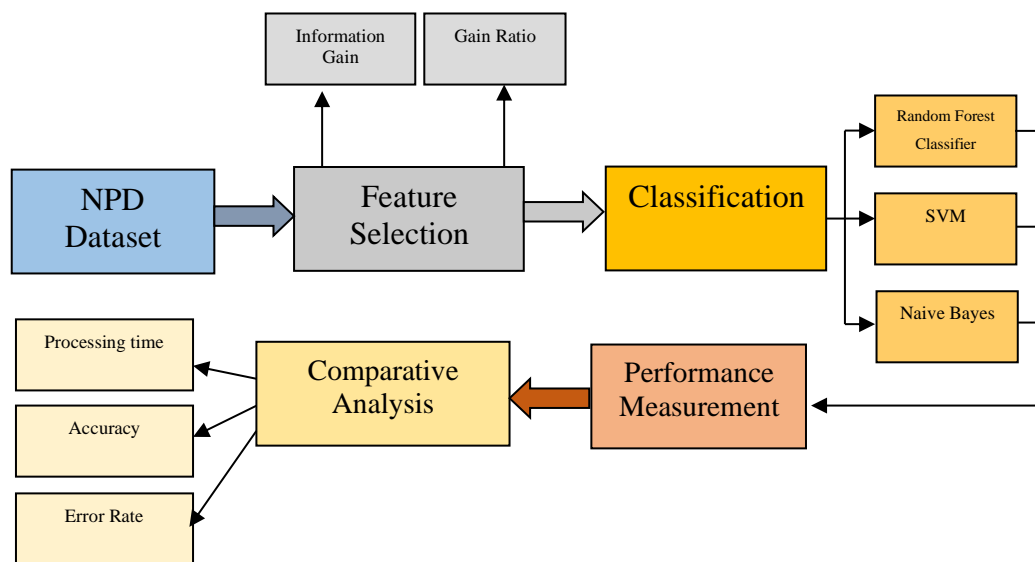
Pada penelitian ini dilakukan analisis SLR terhadap 22 artikel yang terindeks *Scopus Quartile* Q1-Q4. Hal ini dikarenakan artikel yang terindeks *Scopus Quartile* Q1-Q4 memiliki dampak yang signifikan pada kutipan penelitian (Setyaningsih *et al.* 2018) dibandingkan dengan artikel yang tidak terindeks,

dapat dilakukan analisis sitasi (Falagas *et al.* 2008) dan yang umum digunakan pada analisis bibliometrik (Waltman dan Noyons 2018).

Analisis bibliometrik merupakan aplikasi metode statistik dan matematika terhadap literatur seperti buku, majalah, publikasi online, serta media komunikasi lainnya. Pemetaan yang diperoleh dari aplikasi VOSviewer nantinya dapat dijadikan sebagai acuan dalam melakukan analisis konten secara akurat berdasarkan nama peneliti, tahun publikasi, produktivitas peneliti, dan tren riset arsitektur informasi. Pada penelitian ini, analisis bibliometrik dilakukan untuk menganalisis hubungan bibliometrik berdasarkan kata kunci (*co-occurrence*). Selain itu, penulis juga melakukan analisis *publisher publisher* yang menerbitkan artikel terbanyak dan melakukan analisis sitasi.

### **3.2.2. Diagram Alir Penelitian Proses Evaluasi Kinerja Teknik Pemilihan Fitur pada Berbagai Metode *Machine Learning***

Bagian ini menjelaskan tentang langkah-langkah yang dilakukan dalam kegiatan penelitian untuk proses evaluasi kinerja teknik pemilihan fitur pada berbagai metode *machine learning*. Tahap penelitian diawali dengan penentuan masalah, dilanjutkan dengan pengumpulan data. Sebelum data dimasukkan ke dalam model, perlu dilakukan *pre-processing* terhadap data tersebut. Pada tahap ini, data dibersihkan dan disesuaikan untuk diproses pada langkah selanjutnya (pemilihan fitur menggunakan *information gain* dan *gain ratio*). Tahap selanjutnya adalah fitur yang terpilih akan diolah menjadi metode *machine learning* ((*Support Vector Machine* (SVM), *Random Forest Classifier* (RFC), dan *Naive Bayes*). Tahapan penelitian dapat dilihat pada Gambar 6.



Gambar 6. Diagram Alir Proses Evaluasi Kinerja Teknik Pemilihan Fitur

#### a) Dataset

Data yang digunakan dalam penelitian ini diperoleh dari *Open Psychometrics* (<https://www.kaggle.com/datasets>). Data diseleksi berdasarkan usia yaitu digunakan rentang usia 14 – 50 tahun, sehingga data yang terkumpul berjumlah 8.340 data yang terdiri dari 44 fitur. Rentang usia tersebut merupakan usia terbanyak yang mengalami gangguan mental (Ananda 2021). Dalam penelitian ini dilakukan tiga tahapan utama yaitu pemilihan fitur, klasifikasi dan pengukuran kinerja dari masing-masing metode *machine learning*. Tampilan dataset awal yang diperoleh dapat dilihat pada Gambar 7.

score,"Q1","Q2","Q3","Q4","Q5","Q6","Q7","Q8","Q9","Q10","Q11","Q12","Q13","Q14"	
18,2,2,2,2,1,2,1,2,2,2,1,1,2,1,1,1,1,1,2,2,2,1,2,2,2,1,2,1,1,1,2,2,2,1,2,211,1,50	
6,2,2,2,1,2,2,1,2,1,1,2,2,2,1,2,2,1,1,2,1,2,2,1,2,2,2,1,2,2,2,1,2,2,2,2,1,149,1,40	
27,1,2,2,1,2,1,2,1,2,2,2,1,1,1,1,2,2,1,1,2,2,2,1,2,1,1,2,1,2,2,1,1,2,1,1,2,1,2,168,1,28	
29,1,1,2,2,2,1,2,1,1,2,1,1,1,1,1,1,1,2,2,1,2,1,1,1,2,1,2,1,2,2,1,1,2,1,1,2,1,2,2,1,1,230,1,37	
6,1,2,1,1,1,2,1,2,1,2,2,2,2,2,1,1,1,1,1,1,1,2,1,2,2,1,2,1,2,2,2,1,2,2,2,0,1,389,1,50	
19,1,2,2,1,2,1,1,1,2,2,1,1,1,2,1,1,1,1,1,1,1,1,1,1,2,1,1,1,2,1,1,2,1,1,2,1,1,2,2,2,2,361,1,27	
3,1,2,2,1,1,1,1,2,1,1,2,2,2,2,1,2,1,1,1,1,2,1,1,2,2,2,1,2,2,2,1,2,2,1,2,2,2,2,1,388,1,45	
7,2,2,2,1,2,2,2,2,1,1,2,2,2,2,1,1,1,1,1,1,2,1,1,2,2,1,2,2,2,1,2,1,1,2,2,2,1,1,367,2,36	
20,1,2,2,1,1,2,2,1,1,1,1,1,1,1,2,2,1,1,2,2,1,1,2,1,2,1,2,2,2,1,1,1,1,2,1,1,2,1,2,2,1,263,2,24	
23,1,2,1,1,1,1,1,1,1,2,1,1,1,1,1,1,1,2,1,2,1,2,2,1,1,2,2,1,1,1,1,1,1,1,1,2,1,1,2,1,224,2,20	
6,2,2,2,1,2,1,1,2,1,1,2,2,2,1,1,1,1,1,1,1,2,1,1,2,2,2,1,2,2,2,1,2,2,2,1,2,2,1,220,1,25	
35,1,1,1,2,2,1,2,1,2,2,1,1,1,1,2,1,1,1,2,2,1,1,2,1,2,2,1,2,1,1,2,1,1,2,1,1,2,1,1,2,791,1,60	
30,1,1,1,2,1,2,2,1,2,1,1,1,1,1,2,2,1,2,2,2,1,2,1,1,1,2,1,1,1,1,1,2,2,1,1,2,1,1,1,2,262,2,18	
8,1,2,2,1,1,2,1,2,2,2,2,2,2,2,2,1,1,1,2,1,2,1,1,2,2,1,2,2,2,2,1,2,2,1,2,2,2,2,2,3830,1,43	
13,1,2,2,1,1,2,2,2,2,2,2,2,2,2,2,1,2,1,2,1,1,2,2,2,1,2,1,2,1,1,2,1,1,2,2,2,2,246,2,31	
34,1,2,2,2,2,1,2,1,2,2,1,1,1,1,1,1,2,2,1,2,1,1,2,1,2,2,1,2,1,1,1,2,1,1,2,1,1,1,1,2,234,1,48	
12,1,2,2,1,2,2,1,2,1,2,2,2,2,2,1,1,1,1,2,1,2,2,1,2,2,2,1,1,2,1,1,2,1,2,2,2,2,2,244,1,36	
2.2.2.2.1.1.2.1.2.1.1.2.2.2.2.1.2.1.1.1.2.1.1.2.2.2.1.2.2.1.1.2.2.1.2.2.2.1.178.1.52	

Gambar 7. Dataset dengan format .csv

Gambar 7 menjelaskan bahwa terdapat 44 fitur yang terdiri dari fitur score, age, gender, elapse, question 1 – question 4. Dataset berisi 40 pernyataan yang diisi oleh responden dengan memilih jawaban 1 atau 2. Daftar pernyataan tersebut dapat dilihat pada Gambar 8.

Q1. 1= I have a natural talent for influencing people. 2=I am not good at influencing people.  
 Q2. 1=Modesty doesn't become me. 2=I am essentially a modest person.  
 Q3. 1=I would do almost anything on a dare. 2=I tend to be a fairly cautious person.  
 Q4. 1=When people compliment me I sometimes get embarrassed. 2=I know that I am good because everybody keeps telling me so.  
 Q5. 1=The thought of ruling the world frightens the hell out of me. 2=If I ruled the world it would be a better place.  
 Q6. 1=I can usually talk my way out of anything. 2=I try to accept the consequences of my behavior.  
 Q7. 1=I prefer to blend in with the crowd. 2=I like to be the center of attention.  
 Q8. 1=I will be a success. 2=I am not too concerned about success.  
 Q9. 1=I am no better or worse than most people. 2=I think I am a special person.  
 Q10. 1=I am not sure if I would make a good leader. 2=I see myself as a good leader.  
 Q11. 1=I am assertive. 2=I wish I were more assertive.  
 Q12. 1=I like to have authority over other people. 2=I don't mind following orders.  
 Q13. 1=I find it easy to manipulate people. 2=I don't like it when I find myself manipulating people.  
 Q14. 1=I insist upon getting the respect that is due me. 2=I usually get the respect that I deserve.  
 Q15. 1=I don't particularly like to show off my body. 2=I like to show off my body.  
 Q16. 1=I can read people like a book. 2=People are sometimes hard to understand.  
 Q17. 1=If I feel competent I am willing to take responsibility for making decisions. 2=I like to take responsibility for making decisions.  
 Q18. 1=I just want to be reasonably happy. 2=I want to amount to something in the eyes of the world.  
 Q19. 1=My body is nothing special. 2=I like to look at my body.  
 Q20. 1=I try not to be a show off. 2=I will usually show off if I get the chance.  
 Q21. 1=I always know what I am doing. 2=Sometimes I am not sure of what I am doing.  
 Q22. 1=I sometimes depend on people to get things done. 2=I rarely depend on anyone else to get things done.  
 Q23. 1=Sometimes I tell good stories. 2=Everybody likes to hear my stories.  
 Q24. 1=I expect a great deal from other people. 2=I like to do things for other people.  
 Q25. 1=I will never be satisfied until I get all that I deserve. 2=I take my satisfactions as they come.  
 Q26. 1=Compliments embarrass me. 2=I like to be complimented.  
 Q27. 1=I have a strong will to power. 2=Power for its own sake doesn't interest me.  
 Q28. 1=I don't care about new fads and fashions. 2=I like to start new fads and fashions.  
 Q29. 1=I like to look at myself in the mirror. 2=I am not particularly interested in looking at myself in the mirror.  
 Q30. 1=I really like to be the center of attention. 2=It makes me uncomfortable to be the center of attention.  
 Q31. 1=I can live my life in any way I want to. 2=People can't always live their lives in terms of what they want.  
 Q32. 1=Being an authority doesn't mean that much to me. 2=People always seem to recognize my authority.  
 Q33. 1=I would prefer to be a leader. 2=It makes little difference to me whether I am a leader or not.  
 Q34. 1=I am going to be a great person. 2=I hope I am going to be successful.  
 Q35. 1=People sometimes believe what I tell them. 2=I can make anybody believe anything I want them to.  
 Q36. 1=I am a born leader. 2=Leadership is a quality that takes a long time to develop.  
 Q37. 1=I wish somebody would someday write my biography. 2=I don't like people to pry into my life for any reason.  
 Q38. 1=I get upset when people don't notice how I look when I go out in public. 2=I don't mind blending into the crowd when I go out in public.  
 Q39. 1=I am more capable than other people. 2=There is a lot that I can learn from other people.  
 Q40. 1=I am much like everybody else. 2=I am an extraordinary person.

Gambar 8. Daftar pernyataan pada kuesioner

Deskripsi dari masing-masing fitur dapat dilihat pada Tabel 6.

Tabel 6. Deskripsi fitur

Kode Fitur	Fitur	Deskripsi
F1	Jenis Kelamin ( <i>gender</i> )	Jenis kelamin pada dataset awal yaitu male, female, lainnya dan tidak memilih satupun.
F2	Usia ( <i>age</i> )	Usia yang mengisi kuesioner yaitu minimal usia 14 tahun. Sedangkan usia maksimal yang ditemukan pada data awal yaitu 190 tahun.
F3 – F42	Pernyataan 1 – 40 ( <i>question 1 – 40</i> )	Jawaban responden berupa nilai 1 atau 2 sesuai dengan kepribadian responden



Kode Fitur	Fitur	Deskripsi
F43	Elapse	Merupakan selish waktu <i>submitted</i> dengan <i>loaded</i> dalam satuan detik
F44	Skor ( <i>score</i> )	Penjumlahan hasil pengisian kuesioner responden
F45	Kelas	Klasifikasi hasil pengisian kuesiner berupa “Yes” dan “No”

#### b) Pemilihan Fitur

Pemilihan fitur merupakan sebuah teknik untuk memilih fitur pada sebuah data yang tidak memiliki relevansi maupun berlebihan pada sekumpulan data, teknik seleksi fitur bekerja dengan cara menghapus atau membuang data yang tidak relevan atau berlebihan. Tujuan utama dari pemilihan fitur adalah untuk menghindari *overfitting* sehingga meningkatkan performa dari sistem dan sistem menghasilkan hasil yang lebih cepat dan efektif. Dataset yang diperoleh biasanya memiliki banyak fitur dan memiliki fitur yang tidak penting untuk kategorisasi (Calik *et al.* 2019; Yim *et al.* 2020). Fitur yang tidak relevan akan berpengaruh terhadap keakuratan metode *machine learning*. Untuk itu diperlukan strategi pemilihan fitur yang digunakan untuk mengatasi hal ini dan mengurangi dimensi fitur. Pemilihan fitur merupakan proses memilih subset yang relevan dan informatif dari semua fitur yang tersedia dalam data untuk digunakan dalam pelatihan model *machine learning*. Tujuan dari pemilihan fitur adalah untuk meningkatkan kinerja dan efisiensi model dengan mengurangi dimensi data, menghilangkan fitur yang tidak penting atau redundan, serta meningkatkan interpretabilitas dan generalisasi model. Pemilihan fitur dalam *machine learning* memainkan peran penting dalam meningkatkan kinerja model, mengurangi dimensi data, dan meningkatkan interpretabilitas. Namun, pemilihan fitur harus dilakukan dengan hati-hati dan berdasarkan pemahaman domain serta analisis yang tepat.

Terdapat beberapa metode dalam pemilihan fitur yaitu metode *filter*, *wrapper*, *embedded*, *principal component analysis*, *genetic algorithms backward elimination* dan *forward selection*. Setiap teknik memiliki kelebihan dan kelemahan tertentu dan dapat diterapkan tergantung pada sifat data dan tujuan pemilihan fitur. Pemilihan fitur yang baik memerlukan pemahaman yang baik tentang data dan masalah yang dihadapi, serta percobaan yang berulang untuk mengevaluasi kinerja model dengan subset fitur yang berbeda. Berikut beberapa teknik pemilihan fitur:

a) Metode *Filter*

Metode *filter* melakukan pemilihan fitur secara independen dari model *machine learning* yang akan digunakan. Metode ini mengukur hubungan antara setiap fitur dengan target menggunakan metrik statistik atau informasi. Contohnya adalah analisis korelasi, uji t, *chi-square*, atau *mutual information*. Fitur-fitur dengan skor tinggi dianggap penting dan dipilih untuk digunakan dalam model.

b) Metode *Wrapper*

Metode *wrapper* memilih fitur berdasarkan kinerja model yang dibangun dengan subset fitur tertentu. Ini berarti fitur-fitur dievaluasi menggunakan model spesifik dan metrik kinerja seperti akurasi atau presisi. Metode ini memerlukan iterasi yang lebih banyak karena melibatkan pembangunan dan evaluasi model berulang dengan kombinasi fitur yang berbeda. Contohnya adalah *Recursive Feature Elimination* (RFE) dan *forward/backward stepwise selection*.

c) Metode *Embedded*

Metode *embedded* melakukan pemilihan fitur secara terintegrasi dengan proses pembelajaran model. Fitur-fitur dipilih atau diberi bobot berdasarkan kontribusinya terhadap model yang sedang dilatih. Contoh dari metode ini adalah L1 *regularization* (Lasso) dan L2 *regularization* (Ridge) yang memberikan bobot nol pada beberapa fitur, sehingga memungkinkan pemilihan fitur secara otomatis.

d) *Principal Component Analysis (PCA)*

PCA menggabungkan fitur-fitur yang saling berkorelasi menjadi sejumlah komponen utama yang tidak berkorelasi. Komponen-komponen ini dipesan berdasarkan penjelasan varians dari data. Dengan memilih komponen utama dengan varians yang signifikan, pemilihan fitur dapat dilakukan secara tidak langsung.

e) *Genetic Algorithms*

Algoritma genetik adalah pendekatan yang menggunakan konsep seleksi alam untuk memilih fitur-fitur yang optimal. Populasi fitur-fitur diperlakukan sebagai individu dalam populasi genetik. Melalui iterasi generasi, fitur-fitur yang memberikan kinerja terbaik dipertahankan dan dievolusi secara selektif. Metode ini memungkinkan pencarian ruang fitur yang lebih luas, tetapi juga memerlukan komputasi yang lebih mahal.

f) *Backward Elimination dan Forward Selection*

Metode ini merupakan pendekatan yang sederhana tetapi efektif. *Backward elimination* dimulai dengan semua fitur dan secara iteratif menghapus fitur yang paling tidak signifikan berdasarkan metrik kinerja. Di sisi lain, *forward selection* dimulai dengan model kosong dan secara iteratif menambahkan fitur yang memberikan peningkatan signifikan dalam kinerja.

Metode yang paling sering digunakan untuk seleksi fitur adalah metode *filter* (Das *et al.* 2017). Metode *filter* adalah pendekatan yang cepat dan efisien karena tidak melibatkan pelatihan model. Metode ini mengukur hubungan antara setiap fitur dengan target atau output yang ingin diprediksi menggunakan metrik statistik atau informasi. Fitur-fitur dievaluasi secara individual dan diberi skor berdasarkan ukuran tersebut. Fitur-fitur dengan skor tinggi dianggap penting dan dipilih untuk digunakan dalam model. Berikut ini adalah beberapa metode filter yang umum digunakan dalam pemilihan fitur:

a) Analisis Korelasi

Metode ini mengukur tingkat hubungan antara setiap fitur dengan target atau dengan fitur lain dalam dataset. Biasanya, korelasi Pearson digunakan untuk data numerik, sedangkan korelasi kendall atau spearman digunakan untuk data

ordinal atau kategori. Fitur-fitur dengan korelasi yang tinggi terhadap target dipilih karena diyakini memiliki pengaruh yang signifikan pada prediksi.

b) Uji Statistik

Metode ini menggunakan teknik statistik untuk menguji apakah ada perbedaan signifikan antara kelompok data yang berbeda berdasarkan fitur tertentu. Misalnya, uji  $t$  dapat digunakan untuk membandingkan rata-rata fitur antara dua kelompok yang berbeda (misalnya, pasien dengan dan tanpa penyakit). Jika fitur memiliki perbedaan yang signifikan antara kelompok, maka fitur tersebut dianggap penting dalam pemodelan.

c) *Chi-Square*

Metode ini digunakan untuk menentukan apakah ada hubungan signifikan antara fitur kategorikal (nominal) dan target. Misalnya, jika kita memiliki fitur yang mewakili kategori pekerjaan dan ingin melihat apakah ada keterkaitan antara pekerjaan dan risiko penyakit tertentu, uji chi-square dapat memberikan informasi tentang signifikansi hubungan antara kedua variabel tersebut.

d) *Information Gain*

Metode ini digunakan untuk mengukur informasi yang diberikan oleh fitur terhadap pemisahan kelas target. *Information Gain* menghitung pengurangan ketidakpastian (*entropy*) dalam pemisahan kelas target ketika fitur ditambahkan ke dalam pertimbangan. Fitur dengan *Information Gain* yang tinggi dianggap memberikan informasi yang signifikan dan dipilih untuk digunakan dalam model.

e) *Mutual Information*

Metode ini mengukur sejauh mana dua variabel, seperti fitur dan target, saling bergantung atau memberikan informasi satu sama lain. *Mutual Information* mengukur jumlah informasi yang dibagikan antara dua variabel dan dapat digunakan untuk mengevaluasi pentingnya fitur dalam hubungannya dengan target.

f) *Variance Threshold*

Metode ini menghilangkan fitur-fitur dengan varians yang rendah. Fitur-fitur yang memiliki varians yang rendah cenderung memiliki sedikit variasi dalam nilainya dan tidak memberikan banyak informasi dalam membedakan kelas

target. Oleh karena itu, fitur-fitur dengan varians di bawah ambang batas tertentu dihapus dari pemodelan.

Metode *filter* yang populer yaitu *information gain* dan *gain ratio* (Prasetyowati et al. 2021). Batas nilai yang dikenal sebagai *threshold (cutoff)* digunakan untuk menghitung, memilih dan menentukan setiap fitur dalam kumpulan data. Penelitian ini menggunakan *threshold*  $>0.05$ .

### 1) *Information Gain*

*Information Gain* adalah salah satu metode *filter* yang populer dalam pemilihan fitur, terutama dalam konteks klasifikasi. *Information gain* merupakan salah satu metode dari seleksi fitur, dalam proses *information gain* fitur akan dilakukan perankingan. Ranking fitur yang terbesar merupakan fitur yang paling relevan dan memiliki koneksi yang kuat dengan kumpulan data yang terkait. Teknik ini meranking fitur dengan menghitung *Entropy* dari salah satu kelas sebelum dan setelah melakukan proses pengamatan terhadap fitur yang ada pada satu data yang sama. *Information Gain* dihitung dengan membandingkan entropi himpunan data sebelum dan setelah pemisahan berdasarkan fitur tertentu. *Information Gain* adalah perbedaan antara entropi himpunan data awal dengan entropi himpunan data setelah pemisahan. Semakin tinggi *Information Gain*, semakin banyak informasi baru yang diperoleh melalui pemisahan fitur.

Teknik *information gain* memiliki waktu yang paling cepat pada proses pemilihan fitur dibandingkan teknik lainnya (Bhat dan Dutta 2022). Setelah melakukan proses perhitungan *information gain* maka akan diperoleh fitur-fitur dengan nilai *Gain*-nya masing-masing. Fitur dengan nilai *gain* tinggi dan sesuai dengan *threshold* yang diberikan, maka akan menjadi fitur data latih yang baru untuk diproses pada metode *machine learning*. Dalam praktiknya, *Information Gain* digunakan untuk memilih fitur-fitur yang memberikan pemisahan kelas target yang paling signifikan. Fitur dengan *Information Gain* yang tinggi dianggap memberikan informasi yang lebih relevan dalam memprediksi kelas target. Tahapan yang dilakukan dalam proses pemilihan fitur menggunakan metode *information gain* adalah sebagai berikut:

a) Persiapan Data

Langkah pertama yang dilakukan adalah mempersiapkan dataset yang akan digunakan untuk pemilihan fitur. Pastikan data telah diolah dan siap untuk digunakan, termasuk pengkodean fitur kategorikal, penanganan *missing values*, dan normalisasi data jika diperlukan.

b) Menghitung Entropi Awal

Hitung entropi himpunan data awal sebelum adanya pemisahan fitur. Untuk setiap kelas target dalam dataset, hitung probabilitas kemunculan kelas tersebut dan gunakan rumus entropi yang telah dijelaskan sebelumnya.

c) Menghitung *Information Gain*

Selanjutnya, hitung *Information Gain* untuk setiap fitur dalam dataset. Untuk setiap fitur, lakukan pemisahan dataset berdasarkan nilai fitur tersebut. Hitung entropi untuk setiap subset dataset yang dihasilkan. Kemudian, hitung *Information Gain* dengan mengurangkan entropi himpunan data awal dengan jumlah entropi dari subset-subset tersebut.

d) Memilih Fitur dengan *Information Gain* Tertinggi

Identifikasi fitur dengan *Information Gain* tertinggi. Fitur-fitur ini dianggap memberikan informasi yang paling relevan dan signifikan dalam memprediksi kelas target. Anda dapat mengurutkan fitur-fitur berdasarkan nilai *Information Gain* dan memilih sejumlah tertentu fitur teratas.

e) Evaluasi Kinerja Model

Setelah memilih fitur-fitur dengan *Information Gain* tertinggi, gunakan subset dataset yang hanya terdiri dari fitur-fitur tersebut untuk melatih model ML. Evaluasi kinerja model menggunakan metrik yang relevan untuk memastikan bahwa pemilihan fitur telah berhasil dan memberikan hasil yang baik.

f) Pemutakhiran dan Iterasi

Terkadang, setelah melatih model dengan subset fitur, perlu dilakukan pembaharuan pemilihan fitur berdasarkan performa model. Proses yang dapat dilakukan adalah dengan mengulangi langkah-langkah di atas dengan subset fitur yang lebih kecil atau menambahkan fitur-fitur baru yang mungkin relevan.

Metode *information gain* dalam pemilihan fitur memiliki beberapa keunggulan dan kelemahan. Keunggulan metode *information gain* antara lain:

a) Mengukur Relevansi Fitur

*Information gain* membantu mengukur seberapa banyak informasi yang diberikan oleh suatu fitur dalam memisahkan kelas target. Fitur-fitur dengan *information gain* tinggi dianggap lebih relevan dalam memprediksi kelas target. Dengan demikian, metode ini dapat membantu memilih fitur-fitur yang paling informatif dan memiliki dampak signifikan pada model *machine learning*.

b) Sederhana dan Mudah diimplementasikan

Perhitungan *information gain* relatif sederhana dan mudah diimplementasikan. Perhitungannya hanya melibatkan perhitungan entropi dan jumlah sampel dalam setiap subset dataset. Hal ini membuat metode *information gain* dapat diterapkan dengan mudah dalam berbagai algoritma pemilihan fitur.

c) Tidak Bergantung pada Skala Fitur

*Information gain* tidak bergantung pada skala fitur. Artinya, metode ini dapat digunakan untuk dataset yang memiliki fitur-fitur dengan skala yang berbeda, seperti fitur kategorikal dan fitur numerik.

d) Efektif untuk Mengidentifikasi Fitur yang Dominan

*Information gain* dapat membantu mengidentifikasi fitur-fitur yang dominan atau yang memiliki pengaruh kuat pada kelas target. Dengan menggunakan metode ini, kita dapat fokus pada fitur-fitur yang memiliki dampak signifikan dalam memprediksi kelas target dan mengabaikan fitur-fitur yang kurang relevan.

Sedangkan kelemahan dari metode *information gain* adalah sebagai berikut:

a) Tidak Memperhitungkan Hubungan Antar Fitur

*Information gain* hanya memperhitungkan relevansi fitur secara individual terhadap kelas target. Metode ini tidak mempertimbangkan hubungan antar fitur. Oleh karena itu, *information gain* tidak dapat menangkap hubungan atau interaksi antar fitur yang mungkin ada dalam dataset.

b) Bias terhadap Fitur dengan Banyak Nilai

Metode *information gain* cenderung memihak pada fitur-fitur dengan banyak nilai atau kategori. Fitur-fitur ini cenderung memiliki *information gain* yang lebih tinggi hanya karena memiliki lebih banyak kemungkinan untuk membagi dataset. Hal ini dapat menghasilkan bias dalam pemilihan fitur terhadap fitur-fitur dengan banyak nilai, yang mungkin tidak secara substansial mempengaruhi kelas target.

c) Tidak Optimal untuk Data Numerik

*Information gain* secara alami lebih cocok untuk fitur-fitur kategorikal daripada fitur-fitur numerik. Ketika menghadapi fitur numerik, perlu dilakukan diskritisasi atau pengelompokan nilai numerik menjadi interval-interval kategori agar metode *information gain* dapat diterapkan. Namun, hal ini dapat mengurangi keakuratan dan keberagaman informasi yang dihasilkan dari fitur numerik.

d) Sensitif terhadap Fitur yang Memiliki Banyak Nilai yang Serupa

Jika ada beberapa fitur dengan banyak nilai yang serupa atau saling terkait, metode *information gain* mungkin tidak efektif dalam memilih fitur yang paling informatif. Ini karena perbedaan antara *information gain* dari fitur-fitur semacam itu mungkin tidak signifikan, sehingga membuat pemilihan fitur menjadi kurang konsisten dan dapat dipengaruhi oleh faktor-faktor kecil.

## 2) *Gain Ratio*

*Gain ratio* merupakan peningkatan dari *information gain* dalam mengoptimalkan nilai yang dinormalisasi untuk sebuah fitur di dalam klasifikasi dan memperhitungkan jumlah nilai yang mungkin dari suatu fitur dan berusaha untuk mengkompensasi efek ini dalam perhitungan pemilihan fitur (Trabelsi *et al.* 2017; Takci 2018). *Gain Ratio* adalah metode yang digunakan dalam pemilihan fitur untuk mengatasi bias terhadap fitur-fitur dengan banyak nilai atau kategori. *Gain ratio* juga dapat memperbaiki data yang tidak stabil, oleh karena itu metode *gain ratio* cenderung cocok pada data numerik dua kelas, sederhana sehingga komputasi lebih cepat (Kurniabudi *et al.* 2021).



Saat memilih fitur, *gain ratio* mempertimbangkan jumlah dan ukuran dataset (Sulistiani dan Tjahyanto 2017). Selain itu, *gain ratio* juga mempertimbangkan *information gain* dari suatu fitur dan membaginya dengan *split information*. *Split information* mengukur seberapa banyak informasi yang dibutuhkan untuk membagi dataset berdasarkan fitur tersebut. Jika suatu fitur memiliki banyak nilai yang mungkin, maka *split information* akan lebih tinggi, menunjukkan bahwa informasi yang diperlukan untuk memisahkan dataset menjadi lebih besar. Berikut adalah langkah-langkah dalam menghitung *gain ratio*:

(a) Menghitung *Information Gain*

Hitung *Information Gain* untuk setiap fitur dalam dataset. Perhitungan ini mengukur seberapa banyak informasi yang diberikan oleh suatu fitur dalam memisahkan kelas target.

(b) Menghitung *Split Information*

Hitung *split information* untuk setiap fitur dalam dataset. *Split information* dihitung berdasarkan distribusi nilai dari fitur tersebut. Jika fitur memiliki banyak nilai yang mungkin, *split information* akan lebih tinggi.

(c) Menghitung *Gain Ratio*

Hitung *Gain Ratio* untuk setiap fitur dengan membagi *Information Gain* dengan *Split Information*.

(d) Memilih Fitur dengan *Gain Ratio* Tertinggi

Identifikasi fitur dengan *Gain Ratio* tertinggi. Fitur-fitur ini dianggap memberikan informasi yang paling relevan dalam memprediksi kelas target dengan memperhitungkan jumlah nilai yang mungkin dari fitur tersebut.

(e) Evaluasi Kinerja Model

Setelah memilih fitur-fitur dengan *Gain Ratio* tertinggi, gunakan subset dataset yang hanya terdiri dari fitur-fitur tersebut untuk melatih model *machine learning*. Evaluasi kinerja model menggunakan metrik yang relevan untuk memastikan bahwa pemilihan fitur telah berhasil dan memberikan hasil yang baik.

Metode *gain ratio* dalam pemilihan fitur memiliki beberapa keunggulan dan kelemahan yang perlu dipertimbangkan. Berikut adalah keunggulan dari metode *gain ratio*:

(a) Mengatasi Bias Fitur dengan Banyak Nilai

Salah satu keunggulan utama *gain ratio* adalah kemampuannya untuk mengatasi bias terhadap fitur-fitur dengan banyak nilai atau kategori. Dengan mempertimbangkan *split information*, metode ini menyeimbangkan pengaruh fitur-fitur dengan jumlah nilai yang berbeda, sehingga tidak terlalu memihak kepada fitur-fitur yang memiliki banyak nilai yang mungkin.

(b) Menghasilkan Hasil yang Stabil

*Gain Ratio* cenderung menghasilkan hasil yang stabil dan konsisten dalam pemilihan fitur. Hal ini karena perhitungan *Gain Ratio* tidak sensitif terhadap perubahan dalam distribusi kelas target atau perubahan kecil dalam data. Oleh karena itu, metode ini dapat memberikan kestabilan dalam pemilihan fitur ketika diterapkan pada dataset yang berbeda atau saat ada variasi dalam dataset.

Sedangkan kelemahan dari metode *gain ratio* antara lain:

(a) Sensitif terhadap Fitur yang Memiliki Banyak Nilai

Meskipun *gain ratio* membantu mengatasi bias terhadap fitur dengan banyak nilai, metode ini masih dapat memberikan kecenderungan untuk memilih fitur-fitur dengan banyak nilai yang mungkin. Jika dataset memiliki fitur-fitur dengan jumlah nilai yang sangat berbeda, *gain ratio* mungkin memberikan bobot yang tidak seimbang pada fitur-fitur tersebut.

(b) Pengaruh *Outlier*

Metode *gain ratio* dapat dipengaruhi oleh adanya *outlier* dalam dataset. *Outlier* dapat mempengaruhi distribusi nilai fitur dan mengubah informasi pemisahan yang dihasilkan. Oleh karena itu, penting untuk mempertimbangkan penanganan *outlier* sebelum menerapkan metode ini.

(c) Komputasi yang Lebih Rumit

Perhitungan *gain ratio* melibatkan perhitungan *information gain* dan *split information*, yang dapat membutuhkan komputasi yang lebih rumit dibandingkan dengan metode *filter* lainnya. Terutama jika dataset memiliki

banyak fitur atau ukuran dataset yang besar, perhitungan ini dapat memakan waktu dan sumber daya komputasi yang signifikan.

(d) Tidak Cocok untuk Data Numerik

*Gain ratio* lebih cocok untuk fitur-fitur kategorikal daripada fitur-fitur numerik. Pada data numerik, perlu dilakukan diskritisasi atau pengelompokan nilai numerik menjadi interval-interval kategori agar metode *gain ratio* dapat diterapkan. Namun, hal ini dapat menyebabkan kehilangan informasi yang mungkin penting dalam data numerik.

**c) Klasifikasi**

Salah satu teknik *machine learning* yang paling populer saat ini yaitu klasifikasi. Klasifikasi adalah menemukan model dimana fungsinya untuk mengkategorikan dan mencari perbedaan kelas pada data, klasifikasi merupakan konsep yang bertujuan untuk memprediksi kelas agar label kelas tersebut diketahui, yang dianalisis sesuai kelompok data latih dimana label kelas sudah diketahui (Han *et al.* 2013). Tiga tahapan utama yang dilakukan pada klasifikasi (Wafiyah *et al.* 2017) antara lain:

1. Perancangan Model

Pada tahapan ini dilakukan penyelesaian masalah dengan mencari solusi berdasarkan data latih sudah diklasifikasi.

2. Implementasi Model

Tahapan ini dilakukan untuk menentukan kelas untuk data uji berdasarkan parameter-parameter dan model yang telah ditentukan pada tahap perancangan

3. Evaluasi Model

Tahapan ini mengevaluasi model yang sudah diimplementasi dan data uji sudah diklasifikasikan sesuai dengan batas-batas model.

Teknik klasifikasi adalah salah satu pendekatan dalam *machine learning* yang digunakan untuk memprediksi atau mengklasifikasikan data ke dalam kategori atau kelas yang telah ditentukan sebelumnya. Dalam konteks prediksi

gangguan kepribadian, teknik klasifikasi dapat digunakan untuk memprediksi apakah seseorang menderita gangguan kepribadian tertentu berdasarkan fitur-fitur yang diamati atau dikumpulkan dari individu tersebut. Berikut adalah beberapa teknik klasifikasi yang umum digunakan dalam prediksi gangguan kepribadian:

a) *Decision Tree* (Pohon Keputusan)

*Decision tree* adalah model klasifikasi yang membagi data ke dalam kelompok berdasarkan serangkaian keputusan yang didasarkan pada fitur-fitur *input*. Setiap *node* dalam pohon keputusan mewakili keputusan yang diambil berdasarkan fitur-fitur tertentu. Pohon ini dapat digunakan untuk menggambarkan aturan yang digunakan dalam memprediksi gangguan kepribadian berdasarkan fitur-fitur tertentu.

b) *Naive Bayes*

*Naive Bayes* adalah metode klasifikasi yang berdasarkan pada teorema *Bayes*. Metode ini mengasumsikan independensi antara fitur-fitur yang diamati. Dalam konteks prediksi gangguan kepribadian, *naive bayes* dapat digunakan untuk menghitung probabilitas bahwa seseorang menderita gangguan kepribadian berdasarkan fitur-fitur yang diamati.

c) *Logistic Regression*

*Logistic regression* adalah metode klasifikasi yang umum digunakan dalam pemodelan prediksi. Metode ini menggunakan fungsi logistik untuk memprediksi probabilitas kejadian suatu kategori berdasarkan fitur-fitur input. Dalam prediksi gangguan kepribadian, *logistic regression* dapat digunakan untuk memodelkan probabilitas seseorang menderita gangguan kepribadian tertentu berdasarkan fitur-fitur yang diamati.

d) *Support Vector Machine* (SVM)

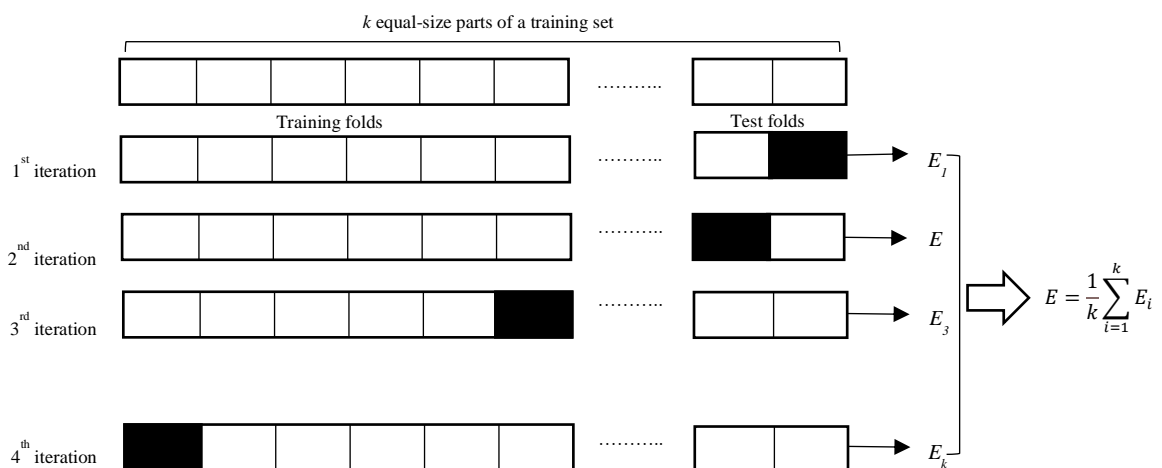
SVM adalah metode klasifikasi yang berfungsi dengan membagi data ke dalam ruang multidimensi dan mencari *hyperplane* yang memaksimalkan jarak antara kelas-kelas yang berbeda. SVM dapat digunakan dalam prediksi gangguan kepribadian dengan mencari *hyperplane* yang dapat memisahkan individu dengan gangguan kepribadian dari individu tanpa gangguan kepribadian berdasarkan fitur-fitur yang diamati.

e) *Random Forest*

*Random Forest* adalah *ensemble learning method* yang menggunakan beberapa pohon keputusan untuk memprediksi kategori target. Setiap pohon dalam *random forest* menghasilkan prediksi dan prediksi dari semua pohon digabungkan untuk menghasilkan hasil akhir. *Random Forest* dapat digunakan dalam prediksi gangguan kepribadian dengan menggabungkan prediksi dari banyak pohon keputusan yang berbeda untuk menghasilkan hasil yang lebih akurat dan stabil.

d) **Pengukuran Kinerja**

Model klasifikasi akan divalidasi menggunakan *k-fold Cross-Validation*. Metode ini umum digunakan untuk *training set* (Yan *et al.* 2022). Prosedur dari *k-fold Cross-Validation* dapat dilihat pada Gambar 9. Sedangkan untuk pengukuran akurasi dan nilai error digunakan *confusion matrix* dengan ukuran  $n \times n$ . Untuk nilai  $n$  menunjukkan jumlah kelas. *confusion matrix*  $n=2$  dapat dilihat pada Tabel 7.



Gambar 9. Prosedur *k-fold Cross-Validation*

Tabel 7. *Confusion matrix*

	<b>Predicted Positives</b>	<b>Predicted Negatives</b>
<b>Actual Positives Instances</b>	Number of True Positives instances (TP)	Number of False Negatives instances (FN)
<b>Actual Negatives Instances</b>	Number of False Positives (FP)	Number of True Negatives instances (TN)

Berdasarkan tabel *confusion matrix* dapat dihitung nilai akurasi dengan menggunakan formula sebagai berikut.

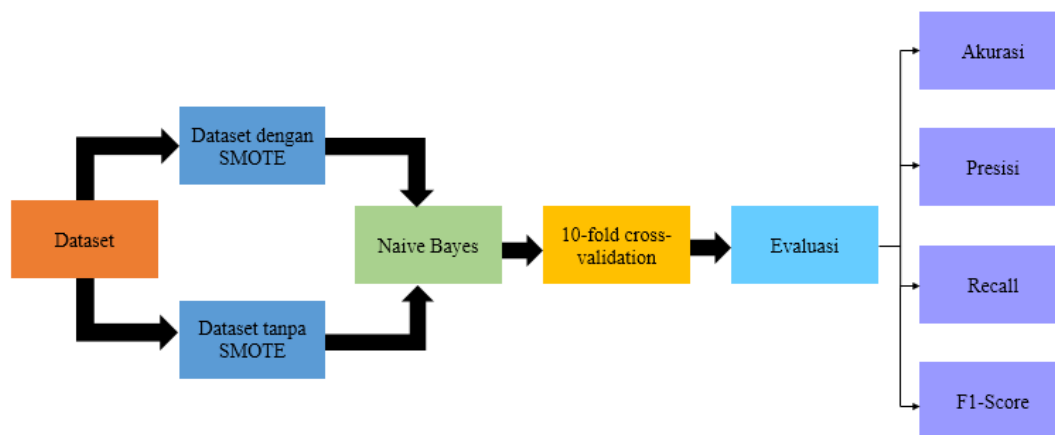
$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

Sedangkan untuk menghitung nilai error dapat digunakan formula sebagai berikut.

$$Missclassification (Error) Rate = \frac{(FP + FN)}{(TP + TN + FP + FN)} \quad (2)$$

### 3.2.3. Diagram Alir Penelitian Implementasi SMOTE-Naive Bayes untuk Mengatasi Ketidakseimbangan Kelas

Bagian ini menjelaskan tentang tahapan yang dilalui pada penelitian. Pengembangan model klasifikasi dalam penelitian ini menggunakan Naive Bayes. Metode ini sudah banyak diterapkan untuk penelitian *machine learning*, salah satu keunggulan dari metode Naive Bayes yaitu algoritma sederhana tetapi mempunyai nilai akurasi yang cukup tinggi (Rachman dan Handayani 2021). Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 10.



Gambar 10. Diagram Alir Penelitian Proses Implementasi SMOTE-Naive Bayes

*Synthetic Minority Over-sampling Technique* (SMOTE) adalah metode *over-sampling* dimana data pada kelas minoritas diperbanyak dengan menggunakan data sintetik yang berasal dari replikasi data pada kelas minoritas. *Over-sampling* pada SMOTE mengambil *instance* dari kelas minoritas lalu mencari *k-nearest neighbor* dari setiap *instance*, kemudian menghasilkan *instance* sintetik daripada mereplikasi *instance* kelas minoritas (Chawla *et al.* 2002). Teknik ini digunakan untuk mengatasi data yang tidak seimbang atau ketidakseimbangan kelas. Ketidakseimbangan kelas terjadi ketika satu kelas dalam dataset memiliki jumlah *instance* yang jauh lebih sedikit dibandingkan kelas lainnya. Ketidakseimbangan ini dapat menyebabkan kinerja model menjadi bias, karena algoritma cenderung lebih condong pada kelas mayoritas dan mengabaikan kelas minoritas (Fitriani *et al.* 2021).

Model yang dibuat dari data yang tidak seimbang akan menghasilkan akurasi prediksi minoritas yang rendah. Informasi yang lebih banyak dari kelas mayoritas akan mendominasi kelas minoritas sehingga menyebabkan batas keputusan yang diragukan dalam sistem klasifikasi (Heranova 2017). Selain itu ketidakseimbangan kelas serta noise juga dapat mempengaruhi kualitas data dalam kinerja klasifikasi. Tahapan yang dilakukan dalam pembuatan data sintetik untuk kelas minoritas menggunakan Teknik SMOTE adalah sebagai berikut:

1. Menginputkan dataset dan jumlah data tambahan yang akan dibuat.

2. Memilih data kelas minoritas. Pada penelitian ini, kelas minoritas yang terdapat pada dataset *narcissistic personality disorder* adalah data kelas negatif.
3. Memisahkan data minoritas (kelas negatif) dan data kelas mayoritas (kelas positif). Setelah data kelas minoritas dan mayoritas dipisahkan dan kemudian menghapus data kelas mayoritas.
4. Pilih dataset minoritas (kelas negatif) secara acak dan hitung data *k-nearest neighbor* yang dipilih. Nilai *k* yang digunakan dalam penelitian ini untuk menghitung *k-nearest neighbor* adalah 3.
5. Setelah itu membuat data baru berdasarkan data yang dipilih secara acak dan *k-nearest neighbor* dengan cara mengalikan jarak yang telah didapatkan pada langkah keempat dengan angka yang dipilih secara acak antara 0 dan 1, kemudian menjumlahkan nilai fitur vektor aslinya.
6. Selanjutnya dilakukan ulang langkah kedua hingga jumlah data baru sesuai dengan jumlah penambahan data yang diinginkan.



## V. KESIMPULAN

Pada bab ini dijelaskan secara singkat dan jelas mengenai kesimpulan dari hasil pengujian dan analisis yang telah dilakukan pada penelitian ini. Bab ini juga disajikan beberapa saran untuk penelitian mendatang.

### 5.1. Kesimpulan

Berdasarkan uraian, penjelasan dan pengujian yang telah dilakukan, didapat beberapa kesimpulan sebagai berikut:

1. Hasil pemrosesan *literature review* menunjukkan bahwa terdapat 22 artikel terindeks Scopus dari 1.022 artikel dari database Google Scholar dan dianalisis menggunakan teknik bibliometrik dan visualisasi menggunakan VOSviewer. Hasil pemrosesan tersebut, ditemukan 138 kata kunci yang dibagi kedalam tujuh kluster. Berdasarkan visualisasi dengan VOSviewer terdapat beberapa topik penelitian yang masih jarang dibahas, antara lain *natural language processing system* untuk gangguan kepribadian, *decision support system*, *data mining*, *deep learning*, *text mining*, *feature selection*.
2. Hasil pemrosesan dalam penentuan subset fitur yang paling terkait dengan label kelas guna meningkatkan kinerja metode *machine learning* dalam menganalisis prediksi *Narcisistic Personality Disorder (NPD)*, penulis mengimplementasikan metode *Information Gain* dan *Gain Ratio*. Pada proses teknik pemilihan fitur digunakan nilai *threshold*  $> 0.05$  dan fitur-fitur akan dirangking berdasarkan nilai *gain* tertinggi. Hasilnya diperoleh informasi bahwa teknik *information gain* menghasilkan fitur yang lebih sedikit dibandingkan teknik *gain ratio*. Teknik *information gain* menghasilkan 37 fitur, sedangkan teknik *gain ratio* menghasilkan 38 fitur. Teknik *information gain* menghapus tujuh fitur yang nilai *threshold*-nya  $< 0.05$ , antara lain f26, f05, f19,

f24, f02, f01 dan f43. Sedangkan dari teknik pemilihan fitur *gain ratio* terdapat enam fitur yang dihapus, antara lain f26, f19, f24, f02, f01 dan f43.

3. Hasil evaluasi kinerja *machine learning* menunjukkan bahwa metode *random forest classifier* telah terbukti lebih baik dalam hal akurasi pada semua skenario pengujian. Nilai akurasi metode *random forest classifier* menggunakan semua fitur, *information gain*, dan *gain ratio* masing-masing sebesar 99,93%, 99,96%, dan 100%. Selain itu diperoleh hasil bahwa metode Naive Bayes dengan menerapkan *information gain* memiliki waktu pemrosesan tercepat; hanya membutuhkan waktu 0,22 detik.
4. Hasil pengolahan data ketidakseimbangan kelas menggunakan teknik SMOTE menunjukkan bahwa teknik tersebut mampu mempengaruhi tingkat akurasi dari algoritma Naive Bayes. Hasil penelitian sebelumnya algoritma Naive Bayes hanya memiliki nilai akurasi sebesar 86,52% sedangkan dengan menerapkan Teknik SMOTE, akurasi algoritma Naive Bayes meningkat menjadi 91,96%.

## 5.2. Saran

Hasil dari penelitian yang telah dilakukan masih perlu dilanjutkan agar menghasilkan penelitian yang lebih baik. Keterbatasan dari penelitian ini antara lain:

1. Hingga saat ini belum ada penelitian yang secara spesifik membahas tentang diagnosa gangguan kepribadian dilengkapi dengan cara pengobatan atau model terapi yang tepat bagi si penderita. Hal ini dikarenakan pengobatan secara dini untuk gejala-gejala gangguan *narcissistic personality disorder* dan gejala-gejala kesehatan mental lainnya telah terbukti dapat mengurangi dampak dan gejala-gejala negatif.
2. Penelitian selanjutnya perlu mempertimbangkan penambahan jumlah dataset dengan penggunaan data baru hasil pengisian kuesioner dan fitur-fitur hasil screening atau wawancara dengan penderita gangguan kepribadian.

3. Metode kecerdasan buatan telah banyak dikenalkan untuk mengatasi kesehatan mental, karena mengklasifikasikan data kesehatan mental merupakan masalah yang sangat menantang, fitur yang digunakan dalam algoritma *machine learning* akan memengaruhi performa klasifikasi secara signifikan. Oleh karena itu, sangat penting bagi peneliti untuk menganalisis data dengan berbagai algoritma *machine learning* untuk memilih akurasi tertinggi di antara algoritma *machine learning* seperti mengevaluasi kinerja menggunakan Teknik *supervised learning* dan *unsupervised learning*.

## VI. DAFTAR PUSTAKA

- Adji TB, R CPB, Permanasari AE. 2015. Expert System for Diagnosis of Personality Disorders with Certainty Factor Approach. Di dalam: *International Conference on Electrical Engineering, Informatics, and Its Education*. hlm. 116–120.
- Al-hajji AA, Alsuhaibani FM, Alharbi NS. 2019. An Online Expert System for Psychiatric Diagnosis. *International Journal of Artificial Intelligence and Applications (IJAIA)*. 10(2):59–76.
- Al-jarrah OY, Yoo PD, Muhaidat S, Karagiannidis GK, Taha K. 2015. Efficient Machine Learning for Big Data : A Review. *Big Data Research*. 2(3):87–93.
- Alsaleh MM, Allery F, Choi JW, Hama T, McQuillin A, Wu H, Thygesen JH. 2023a. Prediction of disease comorbidity using explainable artificial intelligence and machine learning techniques: A systematic review. *Int J Med Inform*. 175(March):1–9.doi:10.1016/j.ijmedinf.2023.105088.
- Alsaleh MM, Allery F, Choi JW, Hama T, McQuillin A, Wu H, Thygesen JH. 2023b. Prediction of disease comorbidity using explainable artificial intelligence and machine learning techniques: A systematic review. *Int J Med Inform*. 175(March):1–9.doi:10.1016/j.ijmedinf.2023.105088.
- Alzahrani S, Bach C. 2014. Impact of Social Media on Personality Development. *International Journal of Innovation and Scientific Research*. 3(2):111–116.
- Alzubi J, Nayyar A, Kumar A. 2018. Machine Learning from Theory to Algorithms : An Overview. Di dalam: *Second National Conference on Computational Intelligence (NCCI 2018)*. hlm. 1–15.
- Ananda SD. 2021. Peran Keluarga dalam Mengatasi Gangguan Kejiwaan bagi Masyarakat Transmigrasi di Desa Harapan Kecamatan Walenrang Kabupaten Luwu.
- Andrew CF, Rogelyn II, Christine. F. Peña, Ken DG. 2018. An Evaluation of SVM and Naive Bayes with SMOTE on Sentiment Analysis Data Set. Di dalam: *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*. hlm. 1–4.
- Ansari N. 2021. A Survey on Feature Selection Techniques using Evolutionary Algorithms. *Iraqi Journal of Science*. 62(8):2796–2812.doi:10.24996/ijs.2021.62.8.32.
- Antosik-wójcinska AZ, Dominiak M, Chojnacka M, Kaczmarek-majer K, Opara KR, Radziszewska W, Olwert A, Ś Ł. 2020. Smartphone as a Monitoring Tool for Bipolar Disorder : a Systematic Review Including Data Analysis ,

- Machine Learning Algorithms and Predictive Modelling. *Int J Med Inform.* 138(November 2019).doi:10.1016/j.ijmedinf.2020.104131.
- Aqilah TM. 2021. Hubungan Kesepian dengan Kecendrungan Perilaku Narsisme Siswa Pengguna Instagram pada Masa Pandemi Covid-19 di Madrasah Aliyah Negeri Kota Baru.
- Aulianto DR, Nashihuddin W. 2020. Bibliometrics and Citation Analysis Of “Baca: Jurnal Dokumentasi dan Informasi” Published During 2015-2019. *Khazanah al-Hikmah : Jurnal Ilmu Perpustakaan, Informasi, dan Kearsipan.* 8(2):149–160.doi:10.24252/kah.v8i2a5.
- Ayuningtyas IPI. 2017. Penerapan strategi penanggulangan penanganan PTSD (Post Traumatic Stress Disorder) pada anak-anak dan remaja. Di dalam: *1st ASEAN School Counselor Conference on Innovation and Creativity in Counseling.* hlm. 47–56.
- Bashir S, Khan ZS, Khan FH, Anjum A, Bashir K. 2019. Improving Heart Disease Prediction Using Feature Selection Approaches. Di dalam: *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST).* IEEE. hlm. 619–623.
- Berdahl CH. 2010. A Neural Network Model of Borderline Personality Disorder. *Neural Networks.* 23(2):177–188.doi:10.1016/j.neunet.2009.10.007.
- Bhat P, Dutta K. 2022. A Multi-tiered Feature Selection Model for Android Malware Detection Based on Feature Discrimination and Information Gain. *Journal of King Saud University - Computer and Information Sciences.* 34(10):1–14.doi:10.1016/j.jksuci.2021.11.004.
- Bickel WK, Tomlinson DC, Craft WH, Ma M, Dwyer CL, Yeh Y-H, Tegge AN, Freitas-Lemos R, Athamneh LN. 2023. Predictors of smoking cessation outcomes identified by machine learning: A systematic review. *Addiction Neuroscience.* 6(August 2022):1–16.doi:10.1016/j.addicn.2023.100068.
- Bilek E, Zang Z, Wolf I, Henrich F, Moessnang C, Braun U, Treede R, Magerl W, Meyer-lindenberG A, Tost H. 2019. Neural network-based alterations during repetitive heat pain stimulation in major depression. *European Neuropsychopharmacology.* 11(23):1–8.doi:10.1016/j.euroneuro.2019.06.011.
- Bilferi Hutapea. 2023. Analisis Pemanfaatan Aplikasi Publish or Perish Terhadap Penulisan Karya Ilmiah Mahasiswa. *PELITA Jurnal Pendidikan dan Keguruan.* 1(1):39–52.
- Budi I, Suryono RR. 2023. Application of named entity recognition method for Indonesian datasets: a review. *Bulletin of Electrical Engineering and Informatics.* 12(2):969–978.doi:10.11591/eei.v12i2.4529.
- Cahyono AS. 2016. Pengaruh Media Sosial Terhadap Perubahan Sosial Masyarakat di Indonesia. *Jurnal Publiciana.* 9(1):140–157.doi:10.32923/asy.v5i2.1586.

- Cai J, Luo J, Wang S, Yang S. 2018. Feature selection in machine learning: A new perspective. *Neurocomputing*. 300(7):70–79. doi:10.1016/j.neucom.2017.11.077.
- Calik N, Belen MA, Mahouti P. 2019. Deep Learning Base Modified MLP Model for Precise Scattering Parameter Prediction of Capacitive Feed Antenna. *Special Issue on Advanced Solution Methods for Modeling Complex Electromagnetic Problems*. 1(1):1–16. doi:10.1002/jnm.2682.
- Chawla N V, Bowyer KW, Hall LO, Kegelmeyer WP. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*. 16(2):321–357.
- Chung J, Teo J. 2022. Mental Health Prediction Using Machine Learning : Taxonomy, Applications, and Challenges. *Applied Computational Intelligence and Soft Computing*. 14(1):1–19. doi:10.1155/2022/9970363.
- Das AK, Goswami S, Chakrabarti A, Chakraborty B. 2017. A New Hybrid Feature Selection Approach using Feature Association Map for Supervised and Unsupervised Classification. *Expert Syst Appl*. 88(12):81–94. doi:10.1016/j.eswa.2017.06.032.
- Dawood S, Schroder HS, Donnellan MB, Pincus AL. 2017. Pathological Narcissism and Nonsuicidal Self-Injury. *J Pers Disord*. 31:1–22.
- Deng GF, Lin WT. 2012. Citation analysis and bibliometric approach for ant colony optimization from 1996 to 2010. *Expert Syst Appl*. 39(6):6229–6237. doi:10.1016/j.eswa.2011.12.001.
- Dewi DSC, Ariana AD. 2021. Pengaruh Kepribadian Narsistik terhadap Perilaku Non-Suicidal Self-Injury pada Remaja Broken Home. *Buletin Riset Psikologi dan Kesehatan Mental*. 1(1):87–95.
- Dey A. 2016. Machine Learning Algorithms : A Review. *International Journal of Computer Science and Information Technologies*. 7(3):1174–1179.
- Dilip V, Jeste, Jeffrey A, Lieberman, David Fassler, Roger Peele. 2013. Diagnostic and Statistical Manual of DSM-5.
- Dipnall JF, Pasco JA, Berk M, Williams LJ, Dodd S, Jacka FN, Meyer D. 2016. Fusing Data Mining, Machine Learning and Traditional Statistics to Detect Biomarkers Associated with Depression. *Data Mine & Machine Learn Biomarkers of Depression*. 11(2):1–23. doi:10.1371/journal.pone.0148195.
- Dwi Fitriani R, Yasin H, Statistika D, Sains dan Matematika F. 2021. Penanganan Klasifikasi Kelas Data Tidak Seimbang dengan Random Oversampling Pada Naive Bayes (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal). *Jurnal Gaussian*. 10(1):11–20.
- Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, Duan Y, Dwivedi R, Edwards J, Eirug A, et al. 2019. Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges , Opportunities , and

- Agenda for Research , Practice and Policy. *Int J Inf Manage.*(August):2–47.doi:10.1016/j.ijinfomgt.2019.08.002.
- van Eck NJ, Waltman L. 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics.* 84(2):523–538.doi:10.1007/s11192-009-0146-3.
- Eck NJ Van, Waltman L. 2014. Measuring Scholarly Impact. Di dalam: *Measuring Scholarly Impact.* hlm. 285–320.
- Ehiabhi J, Wang H. 2023. A Systematic Review of Machine Learning Models in Mental Health Analysis Based on Multi-Channel Multi-Modal Biometric Signals. *BioMedInformatics.* 3(1):193–219.doi:10.3390/biomedinformatics3010014.
- Engkus E, Hikmat H, Saminnurahmat K. 2017. Perilaku Narsis pada Media Sosial di Kalangan Remaja dan Upaya Penanggulangannya. *Jurnal Penelitian Komunikasi.* 20(2):121–134.doi:10.20422/jpk.v20i2.220.
- Faccini L, Allely CS. 2016. Mass violence in individuals with Autism Spectrum Disorder and Narcissistic Personality Disorder: A case analysis of Anders Breivik using the “Path to Intended and Terroristic Violence” model. *Aggress Violent Behav.* 31(1):229–236.doi:10.1016/j.avb.2016.10.002.
- Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. 2008. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal.* 22(2):338–342.doi:10.1096/fj.07-9492lsf.
- Fazel S, O’Reilly L. 2019. Machine Learning for Suicide Research – Can It Improve Risk Factor Identification? *JAMA Psychiatry.* October 23:23–24.doi:10.1002/9780470770771.
- Fernandes BS, Karmakar C, Tamouza R, Tran T, Yearwood J, Hamdani N, Laouamri H, Richard JR, Yolken R, Berk M, *et al.* 2020. Precision psychiatry with immunological and cognitive biomarkers: a multi-domain prediction for the diagnosis of bipolar disorder or schizophrenia using machine learning. *Transl Psychiatry.* 10(1):1–13.doi:10.1038/s41398-020-0836-4.
- Fitriani RD, Yasin H, Tarno. 2021. Penanganan Klasifikasi Kelas Data Tidak Seimbang dengan Random Oversampling pada Naive Bayes (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal). *Jurnal Gaussian.* 10(1):11–20.
- Geneva. 2022. WHO highlights urgent need to transform mental health and mental health care. *WHO int.* [diunduh 2023 Mei 7]. Tersedia pada: <https://www.who.int/news/item/17-06-2022-who-highlights-urgent-need-to-transform-mental-health-and-mental-health-care#:~:text=In 2019%2C nearly a billion,6 years lived with disability.>
- Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, Marsh J, DeVyllder J, Walter M, Berrouiguet S, *et al.* 2021. Machine learning and natural language processing in mental health: Systematic review. *J Med Internet Res.* 23(5):1–20.doi:10.2196/15708.

- Godara S, Singh R. 2016. Evaluation of Predictive Machine Learning Techniques as Expert Systems in Medical Diagnosis. 9(March).doi:10.17485/ijst/2016/v9i10/87212.
- Gowtham. 2021. How I was using Naive Bayes (Incorrectly) till now — Part-2. *Towards Data Science*.
- Hairani H, Saputro KE, Fadli S. 2020. K-means-SMOTE for handling class imbalance in the classification of diabetes with C4.5, SVM, and naive Bayes. *Jurnal Teknologi dan Sistem Komputer*. 8(2):89–93.doi:10.14710/jtsiskom.8.2.2020.89-93.
- Han J, Kamber M, Pei J. 2013. *Data Mining Concept and Techniques*.
- Hartini N, Fardana NA, Ariana AD, Wardana ND. 2018. Stigma Toward People with Mental Health Problems in Indonesia. *Psychol Res Behav Manag*. 11(1):535–541.doi:10.2147/PRBM.S175251.
- Haz L, Rodriguez-Garcia MA, Fernandez A. 2022. Detecting Narcissist Dark Triad Psychological Traits from Twitter. *14th International Conference on Agents and Artificial Intelligence (ICAART) 2022*. 2:313–322.doi:10.5220/0010839100003116.
- He H, Garcia EA. 2009. Learning from imbalanced data. *IEEE Trans Knowl Data Eng*. 21(9):1263–1284.doi:10.1109/TKDE.2008.239.
- Heranova O. 2017. Synthetic Minority Oversampling Technique pada Averaged One Dependence Estimators untuk Klasifikasi Credit Scoring. *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)*. 1(3):443–450.
- Hoffmann K, Doucette L. 2012. A Review of Citation Analysis Methodologies for Collection Management. *Coll Res Libr*. 7(1):321–335.
- Hu Y, Li W, Wright D, Aydin O, Wilson D, Maher O, Raad M. 2019. Artificial Intelligence Approaches. Di dalam: *The Geographic Information Science & Technology Body of Knowledge (3rd Quarter 2019 Edition)*. hlm. 1–12.
- Huang K, Wu C, Su M. 2018. Attention-based Convolutional Neural Network and Long Short-term Memory for Short-term Detection of Mood Disorders based on Elicited Speech Responses. *Attention-based Convolutional Neural Network and Long Short-term Memory for Short-term Detection of Mood Disorders based on Elicited Speech Responses*. Desember.doi:10.1016/j.patcog.2018.12.016.
- Indera Zainul Mutaqien. 2016. Pengembangan Metode Seleksi Fitur dan Tranformasi Data Pada Sistem Deteksi Intrusi dengan Pembatasan Ukuran Cluster dan Sub-Medoid.
- IPKIndonesia IPK (IPK). 2020. Data Keanggotaan.
- Issah I, Appiah O, Appiahene P, Inusah F. 2023. A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. *Decision Analytics Journal*. 7(March):1–11.doi:10.1016/j.dajour.2023.100204.



- Jain T, Jain A, Hada PS, Kumar H, Verma VK, Patni A. 2021. Machine Learning Techniques for Prediction of Mental Health. Di dalam: *The 3rd International Conference on Inventive Research in Computing Applications (ICIRCA) 2021*. IEEE. hlm. 1606–1613.
- Jaiswal S, Song S, Valstar M. 2019. Automatic Prediction of Depression and Anxiety from Behaviour and Personality Attributes. Di dalam: *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE. hlm. 454–460.
- Jan A, Meng H, Member Senior, Falinie Y, Gaus BA, Member Student, Zhang F. 2018. Artificial Intelligent System for Automatic Depression Level Analysis Through Visual and Vocal Expressions. *IEEE Trans Cogn Dev Syst*. 10(3):668–680.doi:10.1109/TCDS.2017.2721552.
- Jan van Eck N, Waltman L. 2022. *VOSviewer Manual*.
- Jan Z, Ai-ansari N, Mousa O, Abd-alrazaq A, Ahmed A, Alam T, Househ M. 2021. The Role of Machine Learning in Diagnosing Bipolar Disorder : Scoping Review. *J Med Internet Res*. 23(11):1–20.doi:10.2196/29749.
- Janssen R, Spronck P, Arntz A. 2014. Case-Based Reasoning for Predicting the Success of Therapy. Di dalam: *Expert System*. Vol. 00. hlm. 1–13.
- Johar A, Hartuti P, Palupi DD. 2014. Implementasi Metode Frame untuk Mendiagnosa Gangguan Kepribadian Dramatik Menggunakan Sistem Pakar. *Jurnal Rekursif*. 2(2):72–80.
- Jung J, Lee H, Jung H, Kim H. 2023. Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: A systematic review. *Heliyon*. 9(5):e16110.doi:10.1016/j.heliyon.2023.e16110.
- Jung Y, Yoon YI. 2017. Multi-level Assessment Model for Wellness Service Based on Human Mental Stress Level. *Multimed Tools Appl*. 76:11305–11317.doi:10.1007/s11042-016-3444-9.
- Kangra K, Singh J. 2023. Comparative analysis of predictive machine learning algorithms for diabetes mellitus. *Bulletin of Electrical Engineering and Informatics*. 12(3):1728–1737.doi:10.11591/eei.v12i3.4412.
- Karim A, Joko Soebagyo, Rahma Puspa Nuranti, Ana Luklu Uljanah. 2021. Analisis Bibliometrik Menggunakan Vosviewer Terhadap Trend Riset Matematika Terapan Di Google Scholar. *Jurnal Riset Pendidikan Matematika Jakarta*. 3(2):23–33.doi:10.21009/jrpmj.v3i2.22264.
- (KemenkesRI) KKRI. 2019. Laporan Riskesdas 2018 Nasional.
- Khan A, Baharudin B, Lee LH, Khan K. 2010. A Review of Machine Learning Algorithms for Text-Documents Classification. *Journal of Advances in Information Technology*. 1(1):4–20.doi:10.4304/jait.1.1.4-20.
- Khazbak M, Wael Z, Ehab Z, Gerorge M, Eliwa E. 2021. MindTime: Deep Learning Approach for Borderline Personality Disorder Detection. 2021

- International Mobile, Intelligent, and Ubiquitous Computing Conference, MIUCC 2021*. 21(01):337–344.doi:10.1109/MIUCC52538.2021.9447620.
- Kim S, Lee HK, Lee K. 2021. Detecting Suicidal Risk using MMPI - 2 Based on Machine Learning Algorithm. *Sci Rep*. 11:1–9.doi:10.1038/s41598-021-94839-5.
- Konna S. 2017. Hubungan Kesehatan Mental dan Healthy Food Choice dengan Kejadian Hipertensi pada Guru Sekolah Menengah di Makassar Tahun 2017.
- Kurniabudi, Harris A, Mintaria AE. 2021. Komparasi Information Gain , Gain Ratio , CFs-Bestfirst dan CFs-PSO Search Terhadap Performa Deteksi Anomali. *Jurnal Media Informatika Budidarma*. 5(1):332–343.doi:10.30865/mib.v5i1.2258.
- Kurniasari NI. 2018. Hubungan antara pola asuh permisif dengan kecenderungan narsistik pada remaja.
- Laijawala V, Achaliya A, Jatta H, Pinjarkar V. 2020. Classification Algorithms based Mental Health Prediction using Data Mining. Di dalam: *Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)*. Vol. 48766. hlm. 1174–1178.
- Lin E, Lin CH, Lane HY. 2020. Precision psychiatry applications with pharmacogenomics: Artificial intelligence and machine learning approaches. *Int J Mol Sci*. 21(3):1–15.doi:10.3390/ijms21030969.
- Liu G, Li Y, Zhang W, Zhang L. 2020. A Brief Review of Artificial Intelligence Applications and Algorithms for Psychiatric Disorders. *Engineering*. 6(4):462–467.doi:10.1016/j.eng.2019.06.008.
- Malhi GS, Bell E. 2020. Prepubertal bipolar disorder: a diagnostic quandary? *Int J Bipolar Disord*. 8(1).doi:10.1186/s40345-020-00187-0.
- Maulana I, Sriati A, Sutini T, Widiati E, Rafiah I, Hidayati NO. 2019. Penyuluhan Kesehatan Jiwa untuk Meningkatkan Pengetahuan Masyarakat tentang Masalah Kesehatan Jiwa di Lingkungan Sekitarnya MKK : Volume 2 No 2 November 2019 Orang yang mengalami gangguan Jiwa di Dunia ini sudah banyak dan bahkan di Indonesia pun banyak p. *Media Karya Kesehatan*. 2(2):218–225.
- Miftahusalam A, Febby Nuraini A, Khoirunisa AA, Pratiwi H. 2020. Perbandingan Algoritma Random Forest, Naive Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer. *Jurnal Informatika Universitas Pamulang*. 5(4):563–584.
- Mijwil MM, Alsaadi A. 2019. Overview of Neural Networks. *Computer Engineering Techniques Department, Baghdad College of Economics Sciences University, Iraq*.(April).

- Mintarya LN, Halim JNM, Angie C, Achmad S, Kurniawan A. 2023. Machine learning approaches in stock market prediction: A systematic literature review. *Procedia Comput Sci.* 216:96–102.doi:10.1016/j.procs.2022.12.115.
- Mitra P, Dimy Fluyau. 2021. Narcissistic Personality Disorder. *National Library of Medicine.*
- Mohri M, Rostamizadeh A, Talwalkar A. 2018. *Foundations of Machine Learning.*
- Mubasyaroh. 2013. Pengenalan Sejak Dini Penderita Mental Disorder. *KONSELING RELIGI: Jurnal Bimbingan Konseling Islam.* 4(1):127–144.
- Nalić J, Martinović G, Žagar D. 2020. New hybrid data mining model for credit scoring based on feature selection algorithm and ensemble classifiers. *Advanced Engineering Informatics.* 45(8):1–9.doi:10.1016/j.aei.2020.101130.
- Nasution K, Huzaeni, Fata A. 2018. Sistem Pakar Diagnosa Penyakit Mata Menggunakan Metode Dempster Shafer Berbasis Web. Di dalam: *Proceeding Seminar Nasional Politeknik Negeri Lhokseumawe.* Vol. 2. hlm. 33–43.
- Noor A. 2019. Sistem Pakar Diagnosa Gangguan Kepribadian. *Jurnal Humaniora Teknologi.* 5(2):33–41.doi:10.34128/jht.v5i2.61.
- Nunes LC, Pinheiro PR, Graduate TCP. 2009. An Expert System Applied to the Diagnosis of Psychological Disorders. Di dalam: *IEEE International Conference on Intelligent Computing and Intelligent Systems.* hlm. 363–367.
- Pahlevi MK, Setiawan BD, Afirianto T. 2018. Identifikasi Gangguan Kepribadian Dramatis Menggunakan Metode Learning Vector Quantization (LVQ). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer.* 2(9):3103–3111.
- Prasetyowati MI, Maulidevi NU, Surendro K. 2021. Determining Threshold Value on Information Gain Feature Selection to Increase Speed and Prediction Accuracy of Random Forest. *J Big Data.* 8(1):1–22.doi:10.1186/s40537-021-00472-4.
- Prasetya J. 2022. Penerapan Klasifikasi Naive Bayes dengan Algoritma Random Oversampling dan Random Undersampling Pada Data Tidak Seimbang Cervical Cancer Risk Factors. *Leibniz : Jurnal Matematika.* 2(2):11–22.
- Primananda AP. 2022. Definisi Mental Illness(Gangguan Mental). *Kementrian Kesehatan Direktorat Jenderal Pelayanan Kesehatan.*:1–3.
- Rachakonda L, Kougianos E. 2018. A Smart Sensor in the IoMT for Stress Level Detection. Di dalam: *2018 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS).* IEEE. hlm. 141–145.
- Rachman A. 2015. Pemanfaatan Media Sosial bagi Penciptaan, Pemeliharaan dan Penyebarluasan Pengetahuan dan Keterampilan Kearifan Lokal di Afrika Timur. *Jurnal Orasi.* VI(1):1–17.

- Rachman A. 2017. Pemanfaatan Media Sosial bagi Penciptaan, Pemeliharaan dan Penyebarluasan Pengetahuan dan Keterampilan Kearifan Lokal di Afrika Timur. *ORASI: Jurnal Dakwah dan Komunikasi*. 6(1):1–17.
- Rachman R, Handayani RN. 2021. Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM. *JURNAL INFORMATIKA*. 8(2):111–122.
- Rahman TG, Ilyas A. 2019. Perilaku narsistik pengguna media sosial di kalangan mahasiswa dan implikasi dalam layanan bimbingan dan konseling narcissistic behavior of social media users in students and implications in guidance and counseling services. *E-Jurnal Pembelajaran Inovasi, jurnal Ilmiah Pendidikan Dasar*. 7(4):1–8.
- Raj S, Masood S. 2020. Analysis and Detection of Autism Spectrum Disorder Using Machine of Learning Analysis and Detection Autism Techniques Spectrum Disorder Using. Di dalam: *International Conference on Computational Intelligence and Data Science (ICCIDS 2019)*. Vol. 167. Elsevier B.V. hlm. 994–1004.
- Randa CPB, Permanasari AE. 2014. Development of Diagnosis Expert System for Personality Disorders. Di dalam: *Makassar International Conference on Electrical Engineering and Informatics (MICEEI)*. hlm. 180–183.
- Reddy US, Thota AV, Dharun A. 2018. Machine Learning Techniques for Stress Prediction in Working Employees. Di dalam: *2018 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2018*. IEEE. hlm. 17–20.
- Rokom. 2021. Kemenkes Beberkan Masalah Permasalahan Kesehatan Jiwa di Indonesia. *Kementrian Kesehatan RI*:1–2.
- Sachithra V, Subhashini LDCS. 2023. How artificial intelligence uses to achieve the agriculture sustainability: Systematic review. *Artificial Intelligence in Agriculture*. 8:46–59.doi:10.1016/j.iaia.2023.04.002.
- Sadya S. 2023. APJII: Pengguna Internet Indonesia 215,63 Juta pada 2022-2023. *DataIndonesia.Id*:1–3.
- Sandhya ND, Raina CK. 2016. A review on Machine Learning Techniques. *International Journal on Recent and Innovation Trends in Computing and Communication*. 4(3):395–399.
- Santi NN. 2017. Dampak Kecenderungan Narsiscisme Terhadap Self Esteem Pada Pengguna Facebook Mahasiswa PGSD UNP. *Jurnal Dimensi Pendidikan dan Pembelajaran*. 5(1):25–30.doi:10.24269/dpp.v5i1.319.
- Saputra AA. 2021. Sistem Pakar Diagnosa Awal Gangguan Kepribadian Narcissistic Personality Disorder.
- Saranya A, Subhashini R. 2023. A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends.

- Decision Analytics Journal*. 7(April):1–14. doi:10.1016/j.dajour.2023.100230.
- Sari MN, Sulyaman, Sulistiono A, Ramadhian MR. 2016. Gangguan Kepribadian dan Perilaku Akibat Penyakit , Kerusakan , dan Personality and Behavioural Disorders due to Disease , Damage , and Brain Dysfunction in A 45 Years Old Men. *J Medula Unila*. 6(1):83–87.
- Sari YS. 2021. Penerapan Metode Naive Bayes Untuk Mengetahui Kualitas Air Di Jakarta. *Jurnal Ilmiah FIFO*. 13(2):222–228. doi:10.22441/fifo.2021.v13i2.010.
- Savitrie E. 2022. Mengenal Pentingnya Kesehatan Mental pada Remaja. *Kementrian Kesehatan Direktorat Jenderal Pelayanan Kesehatan*.:1–3.
- Schlumpf YR, Nijenhuis ERS, Klein C, Jäncke L, Bachmann S. 2019. Functional Reorganization of Neural Networks Involved in Emotion Regulation Following Trauma Therapy for Complex Trauma Disorders. *Neuroimage Clin*. 23(December 2018):1–14. doi:10.1016/j.nicl.2019.101807.
- Setiani Rafika A, Yunan Putri H, Diah Widiarti F, STMIK Raharja Tangerang D, STMIK Raharja Tangerang M, Jendral Sudirman No J. 2017. Analisis Mesin Pencarian Google Scholar Sebagai Sumber Baru untuk Kutipan. *Journal Cerita: Creative Education of Research in Information Technology and Artificial informatics*. 3(2):193–205.
- Setyaningsih I, Indarti N, Jie F. 2018. Bibliometric analysis of the term “green manufacturing.” *International Journal of Management Concepts and Philosophy*. 11(3):315. doi:10.1504/ijmcp.2018.093500.
- Seyedzadeh A, Maroufpoor S, Maroufpoor E, Shiri J. 2019. Artificial Intelligence Approach to Estimate Discharge of Drip Tape Irrigation Based on Temperature and Pressure. *Agric Water Manag*.(November):1–11. doi:10.1016/j.agwat.2019.105905.
- Silvana M, Audina M, Akbar R, Derisma, Firdaus. 2018. Development of Classification Features of Mental Disorder Characteristics Using The Fuzzy Logic Mamdani Method. Di dalam: *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*. IEEE. hlm. 410–414.
- Siringoringo R. 2018. Klasifikasi data tidak Seimbang menggunakan algoritma SMOTE dan k-nearest neighbor. *Jurnal ISD (Journal Information System Development)*. 3(1):44–49.
- Social WA. 2023. Digital 2023 Indonesia. *Digital 2023*.:1–125.
- Society TR. 2017. Machine learning : the power and promise of computers that learn by example.
- Soori M, Arezoo B, Dastres R. 2023. Sustainable Manufacturing and Service Economics. *Sustainable Manufacturing and Service Economics*. 19(1):1–12. doi:10.1016/j.smse.2023.100009.

- Srividya M, Mohanavalli S, Bhalaji N. 2018a. Behavioral Modeling for Mental Health using Machine Learning Algorithms. *J Med Syst.* 42(5):1–12.doi:10.1007/s10916-018-0934-5.
- Srividya M, Mohanavalli S, Bhalaji N. 2018b. Behavioral Modeling for Mental Health using Machine Learning Algorithms. *J Med Syst.* 42(88):1–12.doi:10.1007/s10916-018-0934-5.
- Stolz DS, Vater A, Schott BH, Roepke S, Paulus FM, Krach S. 2021. Reduced frontal cortical tracking of conflict between self-beneficial versus prosocial motives in Narcissistic Personality Disorder. *Neuroimage Clin.* 32(4):1–11.doi:10.1016/j.nicl.2021.102800.
- Sulistiani H, Muludi K, Syarif A. 2021. Implementation of Various Artificial Intelligence Approach for Prediction and Recommendation of Personality Disorder Patient. *J Phys Conf Ser.* 1751(1):1–9.doi:10.1088/1742-6596/1751/1/012040.
- Sulistiani H, Tjahyanto A. 2017. Comparative Analysis of Feature Selection Method to Predict Customer Loyalty. *IPTEK Journal of Engineering.* 3(1):1–5.doi:10.12962/joe.v3i1.2257.
- Suryono RR, Budi I, Purwandari B. 2020. Challenges and trends of financial technology (Fintech): A systematic literature review. *Information (Switzerland).* 11(12):1–20.doi:10.3390/info11120590.
- Sushma SJ, Assegie TA, Vinutha DC, Padmashree S. 2021. An improved feature selection approach for chronic heart disease detection. *Bulletin of Electrical Engineering and Informatics.* 10(6):3501–3506.doi:10.11591/eei.v10i6.3001.
- Syahputra R, Yanris GJ, Irmayani D. 2022. SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter. *Sinkron.* 7(2):671–678.doi:10.33395/sinkron.v7i2.11430.
- Taha A, Cosgrave B, Mckeever S. 2022. Using Feature Selection with Machine Learning for Generation of Insurance Insights. *Applied Sciences.* 12(6):1–16.doi:10.3390/app12063209.
- Takci H. 2018. Improvement of Heart Attack Prediction by the Feature Selection Methods. *Turkish Journal of Electrical Engineering & Computer Sciences.* 26(1):1–10.doi:10.3906/elk-1611-235.
- Teguh D, Fadlil A, Sunardi. 2019. Sistem Pakar Diagnosa Gangguan Kepribadian Menggunakan Metode Dempster Shafer. *Jurnal Sistem Informasi Bisnis.* 01:25–31.
- Teicher MH, Anderson CM, Ohashi K, Khan A, McGreenery CE, Bolger EA, Rohan ML, Vitaliano GD. 2018. Differential effects of childhood neglect and abuse during sensitive exposure periods on male and female hippocampus. *Neuroimage.* 169(4):443–452.doi:10.1016/j.neuroimage.2017.12.055.

- Thieme A, Belgrave D, Doherty G. 2020. Machine Learning in Mental Health: A systematic review of the HCI literature to support the development of effective and implementable ML Systems. *ACM Transactions on Computer-Human Interaction*. 27(5):1–53.doi:10.1145/3398069.
- Trabelsi M, Meddouri N, Maddouri M. 2017. A New Feature Selection Method for Nominal Classifier based on Formal Concept Analysis. Di dalam: *the 21st International Conference on Knowledge - Based and Intelligent Information & Engineering Systems (KES 2017)*. Elsevier B.V. hlm. 186–194.
- Tripathi KP. 2011. A Review on Knowledge-based Expert System : Concept and Architecture. *IJCA Special Issue on “Artificial Intelligence Techniques - Novel Approaches & Practical Applications.”*:19–23.
- Tyagi B, Mishra R, Bajpai N. 2018. Machine Learning Techniques to Predict Autism Spectrum Disorder. *1st International Conference on Data Science and Analytics.*:1–5.
- Utama PKL. 2015. Implementasi Metode Neural Network Pada Perancangan Pengenalan Pola Plat Nomor Kendaraan. Di dalam: *Konferensi Nasional Sistem & Informatika 2015*. hlm. 9–10.
- Victor E, Uyinomen O, Enobakhare E, Festus, A E. 2014. A Fuzzy Inference System for Predicting Depression Risk Levels. *African Journal of Mathematics and Computer Science Research*. 6(January):197–204.doi:10.5897/AJMCSR2013.0511.
- Wafiyah F, Hidayat N, Perdana RS. 2017. Implementasi Algoritma Modified K-Nearest Neighbor ( MKNN ) untuk Klasifikasi Penyakit Demam. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. 1(10):1210–1219.
- Waltman L, Noyons E. 2018. *Bibliometrics for research management and research evaluation: A brief introduction*.
- Wang J, Zhou S, Yi Y, Kong J. 2014. An Improved Feature Selection Based on Effective Range for Classification. *The Scientific World Journal*. 15(1):1–9.doi:10.1155/2014/972125.
- Wardhani KT, Damayanti S, Anggraeny R. 2020. Narcissitic Personality Disorder Pada Tokoh Elena Dalam Komik Tomodachi Gokko Karya Momochi Reiko. *Jurnal SAKURA : Sastra, Bahasa, Kebudayaan dan Pranata Jepang*. 2(1):35–48.doi:10.24843/js.2020.v02.i01.p04.
- Wen C, Liu W, He Z, Liu C. 2023. Research on emergency management of global public health emergencies driven by digital technology: A bibliometric analysis. *Front Public Health*. 10(January):1–16.doi:10.3389/fpubh.2022.1100401.
- Whear R, Bethel A, Abbott R, Rogers M, Orr N, Manzi S, Ukoumunne OC, Stein K, Coon JT. 2022. Systematic reviews of convalescent plasma in COVID-19 continue to be poorly conducted and reported: a systematic review. *J Clin Epidemiol*. 151:53–64.doi:10.1016/j.jclinepi.2022.07.005.

- WHO. 2022. Mental Disorder. *World Health Organization*.:1.
- (WHO) WHO. 2022. Mental Disorders. *WHO.int*.:1.
- Widiyanti W, Solehuddin M, Saomah A. 2017. Profil Perilaku Narsisme Remaja serta Implikasinya Bagi Bimbingan dan Konseling. *Indonesian Journal of Educational Counseling*. 1(1):15–26.doi:10.30653/001.201711.3.
- de Wilde P. 2023. Building performance simulation in the brave new world of artificial intelligence and digital twins: A systematic review. *Energy Build*. 292(April):1–14.doi:10.1016/j.enbuild.2023.113171.
- Yan T, Shen S, Zhou A, Chen X. 2022. Prediction of geological characteristics from shield operational parameters by integrating grid search and K -fold cross validation into stacking classification algorithm. *Journal of Rock Mechanics and Geotechnical Engineering*.:1–12.doi:10.1016/j.jrmge.2022.03.002.
- Yang Z, Ye Q, Chen Q, Ma X, Fu L, Yang G, Yan H, Liu F. 2020. Robust Discriminant Feature Selection Via Joint L<sub>2,1</sub>-Norm Distance Minimization and Maximization. *Knowl Based Syst*. 207(11):1–30.doi:10.1016/j.knosys.2020.106090.
- Yim SJ, Lui LMW, Lee Y, Rosenblat JD, Ragguett R, Park C, Subramaniapillai M, Cao B, Zhou A, Rong C, *et al*. 2020. The utility of smartphone-based , ecological momentary assessment for depressive symptoms. *J Affect Disord*. 274(9):602–609.doi:10.1016/j.jad.2020.05.116.
- Yulia Hayuningtyas R. 2019. Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita. *Jurnal Informatika*. 6(1):18–22.
- Yuwono DT, Fadlil A, Sunardi. 2019. Comparative Analysis of Dempster-Shafer Method and Certainty Factor Method On Personality Disorders Expert Systems. *Scientific Journal of Informatics*. 6(1):12–22.



## LAMPIRAN

- Lampiran 1. Bukti Artikel Telah terbit pada *International Conference on Applied Sciences Mathematics and Informatics (ICASMI 2020)*
- Lampiran 2. Sertifikat sebagai presenter di ICASMI 2020
- Lampiran 3. Bukti Submit pada Jurnal *Bulletin of Electrical Engineering and Informatics (BEEI)*
- Lampiran 4. Bukti sedang dalam proses editing di Jurnal *Bulletin of Electrical Engineering and Informatics (BEEI)*
- Lampiran 5. Bukti artikel telah diterima pada Jurnal *Bulletin of Electrical Engineering and Informatics (BEEI)*
- Lampiran 6. Artikel Penelitian yang Telah Submit di International Journal of Intelligent Systems and Applications in Engineering (IJISAE)