

**TESIS**

**PARTICLE SWARM OPTIMIZATION UNTUK PENENTUAN  
PARAMETER DAN FITUR SELEKSI PADA ALGORITMA  
K-NEAREST NEIGHBOR PADA PREDIKSI TARGET  
HAFALAN SISWA SMPIT MIS**



Disusun oleh:

**Nama : Muhammad Evandry Dewaksana**  
**NIM : 20.55.1341**  
**Konsentrasi : Business Intelligence**

**PROGRAM STUDI S2 TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**TESIS**

**PARTICLE SWARM OPTIMIZATION UNTUK PENENTUAN  
PARAMETER DAN FITUR SELEKSI PADA ALGORITMA K-NEAREST  
NEIGHBOR PADA PREDIKSI TARGET HAFALAN SISWA SMPIT MIS**

**PARTICLE SWARM OPTIMIZATION FOR PARAMETER  
DETERMINATION AND SELECTION FEATURES IN K-NEAREST  
NEIGHBOR ALGORITHM ON THE PREDICTION OF MIS SMPIT  
STUDENTS' MEMORY TARGETS**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

**Nama : Muhammad Evandry Dewaksana**  
**NIM : 20.55.1341**  
**Konsentrasi : Business Intelligence**

**PROGRAM STUDI S2 TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2023**

**HALAMAN PENGESAHAN**

**PARTICLE SWARM OPTIMIZATION UNTUK PENENTUAN PARAMETER  
DAN FITUR SELEKSI PADA ALGORITMA K-NEAREST NEIGHBOR PADA  
PREDIKSI TARGET HAFALAN SISWA SMPIT MIS**

**PARTICLE SWARM OPTIMIZATION FOR PARAMETER DETERMINATION  
AND SELECTION FEATURES IN K-NEAREST NEIGHBOR ALGORITHM ON  
THE PREDICTION OF MIS SMPIT STUDENTS' MEMORY TARGETS**

Dipersiapkan dan Disusun oleh

**Muhammad Evandry Dewaksana**

**20.55.1341**

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis  
Program Studi S2 Teknik Informatika  
Program Pascasarjana Universitas AMIKOM Yogyakarta  
pada **Senin, 3 April 2023**

Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer

Yogyakarta, 3 April 2023

**Rektor**

**Prof. Dr. M. Suyanto, M.M.**  
**NIK. 190302001**

## HALAMAN PERSETUJUAN

**PARTICLE SWARM OPTIMIZATION UNTUK PENENTUAN PARAMETER  
DAN FITUR SELEKSI PADA ALGORITMA K-NEAREST NEIGHBOR PADA  
PREDIKSI TARGET HAFALAN SISWA SMPIT MIS**

**PARTICLE SWARM OPTIMIZATION FOR PARAMETER DETERMINATION  
AND SELECTION FEATURES IN K-NEAREST NEIGHBOR ALGORITHM ON  
THE PREDICTION OF MIS SMPIT STUDENTS' MEMORY TARGETS**

Dipersiapkan dan Disusun oleh

**Muhammad Evandry Dewaksara**  
**20.55.1341**

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis  
Program Studi S2 Teknik Informatika  
Program Pascasarjana Universitas AMIKOM Yogyakarta  
pada **Senin, 3 April 2023**

**Pembimbing Utama**

**Anggota Tim Penguji**

**Prof. Dr. Ema Utami, S.Si., M.Kom.**  
**NIK. 190302037**

**Alva Hendi Muhammad, S.T., M.Eng., Ph.D.**  
**NIK. 190302493**

**Pembimbing Pendamping**

**Dhani Ariatmanto, M.Kom., Ph.D.**  
**NIK. 190302197**

**Muhammad Rudvanto Arief, M. T.**  
**NIK. 190302163**

**Prof. Dr. Ema Utami, S.Si., M.Kom.**  
**NIK. 190302037**

Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer

Yogyakarta, 3 April 2023  
**Direktur Program Pascasarjana**

**Dr. Kusriani, M.Kom.**  
**NIK. 190302106**

## HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Muhammad Evandry Dewaksara**  
NIM : **20.55.1341**  
Konsentrasi : **Business Intelligence**

Menyatakan bahwa Tesis dengan judul berikut:

**Particle Swarm Optimization untuk Penentuan Parameter dan Fitur Seleksi pada Algoritma K-Nearest Neighbor pada Prediksi Target Hafalan Siswa SMPIT MIS**

Dosen Pembimbing Utama : **Prof. Dr. Ema Utami, S.Si., M.Kom.**  
Dosen Pembimbing Pendamping : **M. Rudyanto Arief, M.T.**

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 3 April 2023

Yang Menyatakan,



METERAI TEMPORAL  
10000  
5R2AKG256498

Muhammad Evandry Dewaksara

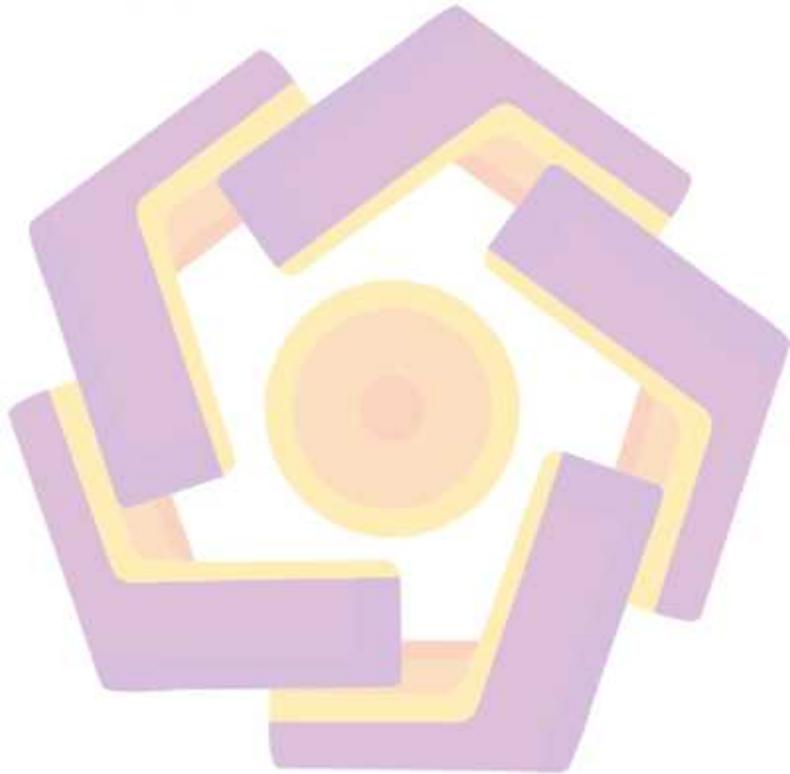
## HALAMAN PERSEMBAHAN

Alhamdulillah, Saya sangat bersyukur kepada Allah Subhanahu Wa Ta'ala karena telah berhasil menyelesaikan tesis ini. Ini semua berkat anugerah dan kasih sayang yang diberikan-Nya, maka penulis mempersembahkan tesis ini kepada semua pihak yang terlibat secara langsung ataupun tidak langsung dalam proses pembuatan tesis.

1. Orang tua dan saudara kakak-kakak, yang selalu mendoakan, memberikan semangat untuk menjalani perkuliahan
2. Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas AMIKOM Yogyakarta yang telah memberikan kesempatan kepada penulis untuk melanjutkan Studi jenjang Strata 2 Program Studi Magister Teknik Informatika di Universitas Amikom Yogyakarta.
3. Prof. Dr. Ema Utami, S.Si., M.Kom dan Bapak M. Rudyanto Arief, M.T. yang telah membimbing penulis dari awal sampai akhir proses pembuatan tesis.
4. Semua pihak yang tidak bisa disebutkan satu persatu yang sudah memberi semua ilmu pengetahuan, informasi dan segalanya sehingga penulis bisa menyelesaikan tesis ini

## HALAMAN MOTTO

"Jika Anda merasa kesepian, ingatlah bahwa Allah Subhanahu Wa Ta'ala mungkin sedang menjauhkan orang-orang dari Anda untuk membuat Anda lebih dekat dengan-Nya."



## KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Allah Subhanahu Wa Ta'ala atas nikmat dan karunia-Nya sehingga penulis dapat menyelesaikan tesis yang berjudul "Particle Swarm Optimization untuk Penentuan Parameter dan Fitur Seleksi pada Algoritma K-Nearest Neighbor pada Prediksi Target Hafalan Siswa SMPIT MIS". Penulis menyadari tesis ini tidak dapat diselesaikan dengan baik tanpa bimbingan, saran dan motivasi dari berbagai pihak.

Peneliti mengucapkan terimakasih yang sebesar-besarnya kepada :

1. Prof. Dr. M. Suyanto, MM, selaku Rektor Universitas AMIKOM Yogyakarta.
2. Prof. Dr. Ema Utami, S.Si., M.Kom. dan Bapak M. Rudyanto Arief, M.T. selaku Dosen Pembimbing.
3. Bapak Alva Hendi Muhammad, S.T., M.Eng., Ph.D., Bapak Dr. Kumara Ari Yuana, S.T., M.T. dan Prof. Dr. Ema Utami, S.Si., M.Kom. selaku Dosen Penguji.
4. Orang tua serta saudara selaku wali yang telah memberikan dukungan dan motivasi.

Semoga Allah Subhanahu Wa Ta'ala memberikan balasan yang lebih kepada pihak yang telah membantu penulis menyelesaikan tesis ini. Semoga tesis ini dapat bermanfaat bagi masyarakat.

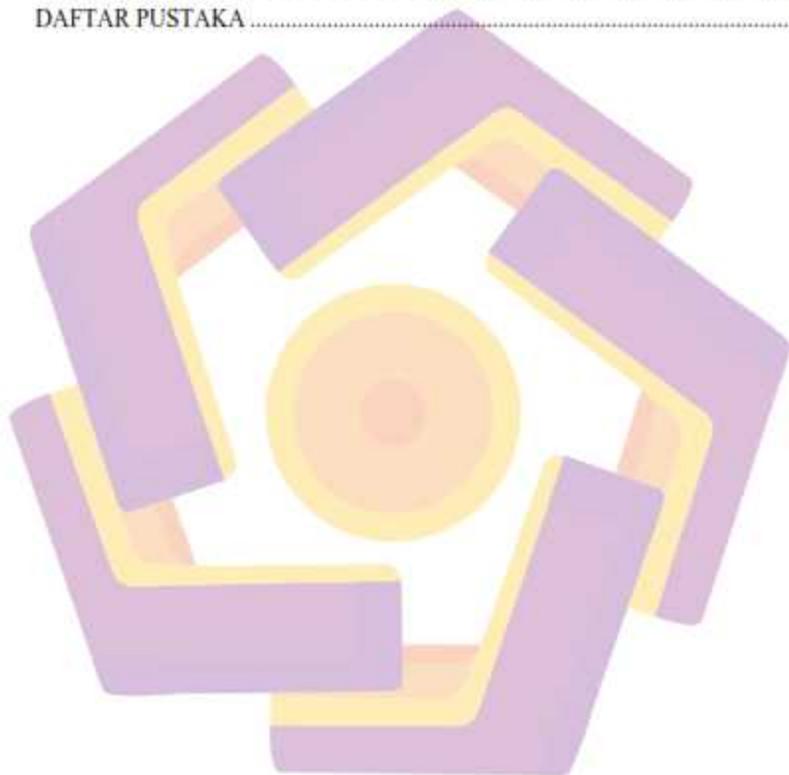
Yogyakarta, 3 April 2023

Penulis

## DAFTAR ISI

HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS .....	v
HALAMAN PERSEMBAHAN .....	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiii
INTISARI.....	xv
<i>ABSTRACT</i> .....	xvi
<b>BAB I PENDAHULUAN</b> .....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	9
1.3. Batasan Masalah.....	9
1.4. Tujuan Penelitian .....	10
1.5. Manfaat Penelitian .....	11
<b>BAB II TINJAUAN PUSTAKA</b> .....	12
2.1. Tinjauan Pustaka.....	12
2.2. Keaslian Penelitian.....	26
2.3. Landasan Teori.....	36
<b>BAB III METODE PENELITIAN</b> .....	47
3.1. Jenis, Sifat dan Pendekatan Penelitian.....	47
3.2. Metode Pengumpulan Data.....	47
3.3. Metode Analisis Data.....	48
3.4. Alur Penelitian .....	48
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN</b> .....	54
4.1. Pengolahan Data.....	54
4.2. Penerapan Algoritma K-Nearest Neighbor, Forward Selection, dan Particle Swarm Optimization (PSO) .....	55
4.2.1. Penerapan Algoritma KNN.....	56

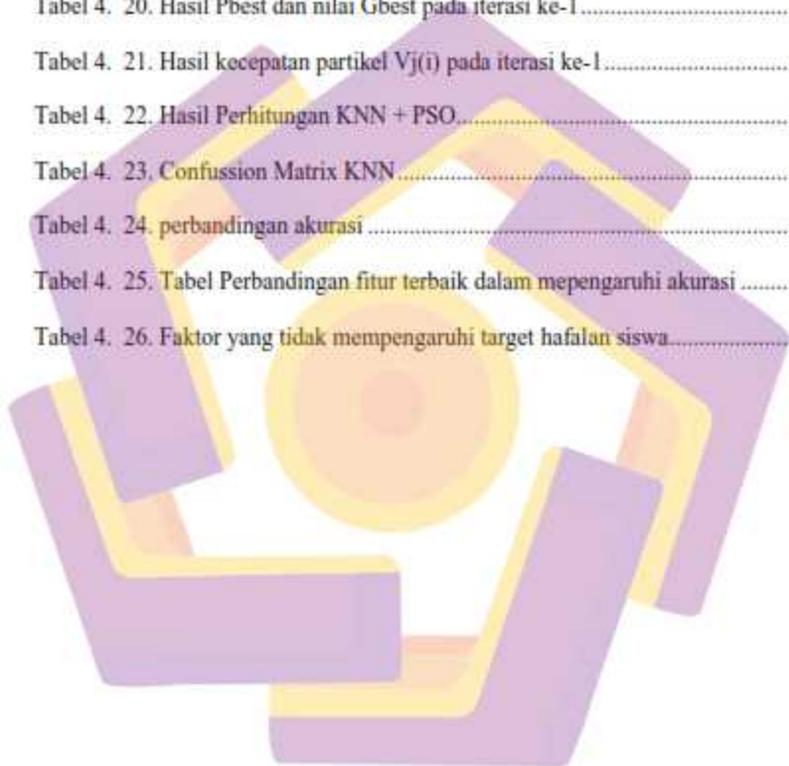
4.2.2. Penerapan Algoritma FS pada KNN.....	64
4.2.3. Penerapan Algoritma PSO pada KNN.....	72
4.3. Hasil Penelitian.....	86
<b>BAB V PENUTUP.....</b>	<b>90</b>
5.1. Kesimpulan.....	90
5.2. Saran.....	91
<b>DAFTAR PUSTAKA.....</b>	<b>93</b>



## DAFTAR TABEL

Tabel 2. 1. Matriks literatur review dan posisi penelitian Particle Swarm Optimization Untuk Penentuan Parameter Dan Fitur Seleksi Pada Algoritma K-Nearest Neighbor Pada Prediksi Target Hafalan Siswa SMPIT MIS .....	26
Tabel 2. 2. Representasi Confusion Matrix .....	46
Tabel 4. 1. Dataset siswa .....	54
Tabel 4. 2. Data latih .....	55
Tabel 4. 3. Data Uji .....	55
Tabel 4. 4. Hasil perhitungan data uji pertama dengan semua data pelatihan .....	57
Tabel 4. 5. Hasil pengurutan perhitungan data uji pertama dengan semua data pelatihan metode KNN .....	57
Tabel 4. 6. Jarak terdekat sesuai nilai K Metode K-NN .....	58
Tabel 4. 7. Label aktual dan label prediksi hasil klasifikasi semua data uji metode K-NN .....	58
Tabel 4. 8. Confussion Matrix KNN .....	59
Tabel 4. 9. Hasil Perhitungan KNN .....	61
Tabel 4. 10. Hasil Perhitungan KNN + Forward Selection .....	69
Tabel 4. 11. Confussion Matrix KNN .....	71
Tabel 4. 12. Simbol-simbol yang digunakan dalam proses perhitungan metode KNN + PSO .....	72
Tabel 4. 13. Tabel jumlah partikel serta nilainya dimasing-masing partikel .....	74
Tabel 4. 14. Nilai error rate di setiap partikel .....	75
Tabel 4. 15. Hasil nilai fitness evaluasi disetiap partikel pada iterasi ke-0 .....	75

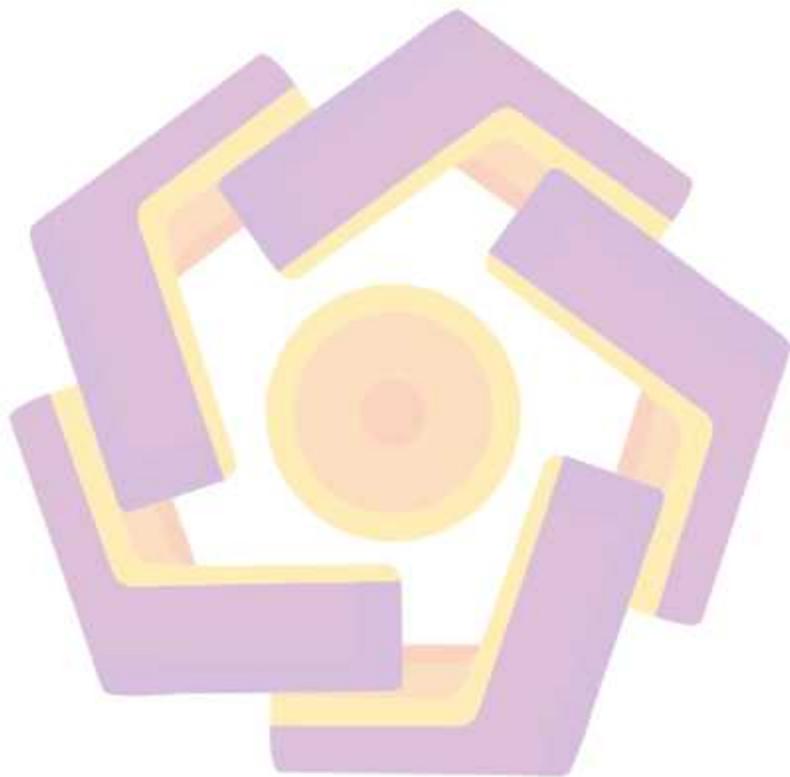
Tabel 4. 16. Hasil Pbest dan Gbest pada iterasi ke-0.....	76
Tabel 4. 17. Hasil kecepatan awal partikel pada iterasi ke-0.....	76
Tabel 4. 18. Hasil nilai posisi partikel pada iterasi ke-1.....	77
Tabel 4. 19. Hasil fitness nilai evaluasi setiap partikel pada iterasi ke-1.....	78
Tabel 4. 20. Hasil Pbest dan nilai Gbest pada iterasi ke-1.....	79
Tabel 4. 21. Hasil kecepatan partikel $V_j(i)$ pada iterasi ke-1.....	80
Tabel 4. 22. Hasil Perhitungan KNN + PSO.....	84
Tabel 4. 23. Confussion Matrix KNN.....	86
Tabel 4. 24. perbandingan akurasi.....	87
Tabel 4. 25. Tabel Perbandingan fitur terbaik dalam mempengaruhi akurasi.....	88
Tabel 4. 26. Faktor yang tidak mempengaruhi target hafalan siswa.....	88



## DAFTAR GAMBAR

Gambar 2. 1. Blok Diagram Model Klasifikasi .....	38
Gambar 2. 2. Flowchart K-Nearest Neighbor.....	40
Gambar 2. 3. Tahapan algoritma Particle Swarm Optimization .....	45
Gambar 3. 1. Diagram alur penelitian bagian 1 .....	49
Gambar 3. 2. Diagram alur penelitian bagian 2 .....	50
Gambar 4. 1. Akurasi K value KNN.....	61
Gambar 4. 2. Metrik evaluasi KNN .....	62
Gambar 4. 3. Menelusuri nama kolom dari dataset .....	65
Gambar 4. 4. Menampilkan semua fitur dan mengecek missing value .....	65
Gambar 4. 5. Perhitungan menggunakan 10 variabel .....	66
Gambar 4. 6. 10 variabel terpilih .....	66
Gambar 4. 7. contoh dataset ketika variabel berjumlah 10.....	66
Gambar 4. 8. akurasi KNN ketika variabel berjumlah 10.....	67
Gambar 4. 9. akurasi KNN ketika variabel berjumlah 11.....	68
Gambar 4. 10. grafik perbandingan akurasi ketika K bernilai 3 .....	69
Gambar 4. 11. Metrik evaluasi KNN + Forward Selection .....	70
Gambar 4. 12. Menerapkan algoritma optimisasi partikel untuk memilih subset fitur terbaik untuk regresi linier.....	73
Gambar 4. 13. Mengatur parameter untuk algoritma optimisasi partikel untuk pemilihan fitur dalam regresi linier .....	73
Gambar 4. 14. source code proses pada PSO.....	81
Gambar 4. 15. PSO Iteration.....	82

Gambar 4. 16. Hasil selection fitur PSO .....	83
Gambar 4. 17. Contoh tampilan dataset .....	83
Gambar 4. 18. Akurasi K value KNN berdasarkan PSO .....	84
Gambar 4. 19. Metrik evaluasi KNN .....	85



## INTISARI

Karantina Al-Qur'an merupakan kegiatan menghafalkan Al-Qur'an dan menanamkan ayat-ayat Allah ke dalam hati dan pikiran sehingga nantinya dapat diucapkan tanpa perlu melihat teksnya. Ada banyak teknik pada proses menghafal ayat suci Al-Qur'an dan terdapat berbagai faktor yang mempengaruhi kelancarannya. SMPIT Makassar Islamic School merupakan Boarding School yang menyediakan kegiatan Karantina Al-Qur'an sebagai salah satu program studinya. Tujuan dari penelitian ini adalah mengetahui faktor apa saja yang mempengaruhi siswa dalam menghafal Al-Qur'an sehingga ketika pembagian kelompok pada program karantina Al-Qur'an, maka pihak sekolah memiliki standar kriteria sebagai acuan.

Pada penelitian ini terdapat 63 siswa yang terdiri dari 37 siswa yang lulus, dan 26 siswa yang tidak lulus. Penelitian ini menggunakan metode deskriptif kuantitatif dengan pendekatan korelasional. Dalam proses klasifikasi, peneliti menggunakan K-Nearest Neighbor(KNN), Forward Selection(FS) dan Particle Swarm Optimization(PSO). FS dan PSO digunakan sebagai seleksi fitur untuk mengoptimalkan jumlah fitur yang digunakan. Perbandingan pada algoritma FS dan PSO, didapatkan hasil bahwa penggunaan fitur yang optimal untuk meningkatkan akurasi KNN yang hanya mendapatkan tingkat akurasi sebesar 69% dengan nilai  $K=3$ , dan nilai akurasi 84% serta  $K=6$  didapatkan ketika menggunakan KNN yang dioptimasi menggunakan PSO, hal serupa juga terjadi ketika KNN dioptimasi menggunakan FS dengan nilai akurasi 84% serta  $K=3$ .

Hasil penelitian dari penelitian ini mendapatkan informasi bahwa faktor-faktor yang mempengaruhi siswa yang lulus dan tidak lulus dalam menghafal Al-Qur'an adalah faktor internal dan eksternal sehingga nantinya pembagian kelompok menghafal bisa lebih seimbang.

Kata Kunci: Hafalan Al-Qur'an, Motivasi, K-Nearest Neighbor, Particle Swarm Optimization, Forward Selection.

## ABSTRACT

*Al-Qur'an quarantine is an activity of memorizing the Al-Qur'an and instilling Allah's verses into the heart and mind so that later it can be recited without the need to look at the text. There are many techniques in the process of memorizing the holy verses of the Qur'an and there are various factors that affect its fluency. SMPIT Makassar Islamic School is a Boarding School that provides Al-Qur'an Quarantine activities as one of its study programs. The purpose of this study was to find out what factors influenced students in memorizing the Al-Qur'an so that when dividing groups into the Al-Qur'an quarantine program, the school had standard criteria as a reference.*

*In this study there were 63 students consisting of 37 students who passed and 26 students who did not pass. This study uses a quantitative descriptive method with a correlational approach. In the classification process, researchers use K-Nearest Neighbor (KNN), Forward Selection (FS) and Particle Swarm Optimization (PSO). FS and PSO are used as feature selection to optimize the number of features used. Comparison of the FS and PSO algorithms, the results show that the optimal use of features to increase the accuracy of KNN only gets an accuracy rate of 69% with a value of  $K = 3$ , and an accuracy value of 84% and  $K = 6$  is obtained when using KNN which is optimized using PSO, the same thing happens when KNN is optimized using FS with an accuracy value of 84% and  $K=3$ .*

*The results of this study obtained information that the factors that influence students who passed and did not pass in memorizing the Qur'an were internal and external factors so that later the distribution of memorization groups could be more balanced.*

*Keywords: Memorizing Al-Qur'an, Motivation, K-Nearest Neighbor, Particle Swarm Optimization, Forward Selection.*

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Menghafal ayat suci Al-Qur'an merupakan aktifitas yang dimana manusia atau khususnya hamba Allah menanamkan ayat-ayat Allah ke dalam hati dan pikirannya sehingga nantinya dapat mengucapkan ayat-ayat tersebut tanpa melihat teks. Ada banyak cara untuk menghafalkan ayat suci Al-Qur'an dan terdapat faktor-faktor yang mempengaruhi prosesnya.

SMPIT Makassar Islamic School merupakan sekolah asrama yang dimana mereka berfokus pada beberapa hal dalam mengajarkan peserta didik diantaranya kemampuan berbahasa, teknologi informasi, olahraga, dan berbagai pelajaran umum lainnya, selain itu akan ada masa karantina Al-Qur'an, yang di mana siswa akan dibebaskan dari semua kegiatan pembelajaran, dan hanya fokus untuk menghafal ayat suci Al-Qur'an. Meski begitu pengelola karantina Al-Qur'an terkadang kebingungan untuk menentukan siswa mana yang masuk kedalam kelompok yang perlu perhatian lebih dan kelompok yang tidak begitu membutuhkan pembinaan lebih lanjut, hal ini didasari oleh kurangnya tenaga pendidik yang berkompeten dalam hal tersebut.

Dalam proses menghafal ayat suci Al-Qur'an, ada faktor-faktor baik dari dalam diri sendiri maupun dari luar yang dapat mempengaruhi seberapa mudah atau sulitnya proses tersebut (Darlimatul, 2008) faktor kepribadian sebagai pembeda dari setiap individu menjadi faktor internal yang paling berpengaruh pada proses menghafal ayat suci Al-Qur'an, kemudian metode menghafal

merupakan faktor eksternal yang berdasarkan prosesnya menjadi faktor seragam bagi setiap individu.

Dalam proses belajar, semuanya diproses dalam bentuk ingatan, sehingga hal-hal yang diingat dapat terikat kuat dalam diri sendiri untuk beradaptasi dengan kehidupan setiap orang. Oleh karena itu, ketika terdapat siswa yang tidak begitu kuat dalam menghafal, namun dengan ketekunannya yang memiliki motivasi yang besar, sehingga dia mampu menghafal dengan sangat baik, maka hal ini dapat dijadikan contoh bagi yang lainnya. (Qardhawi, 2015) Ajaran dari Al-Qur'an dan As-Sunnah melahirkan sebuah landasan dalam pikiran dan jiwa yang di mana akan tumbuh dan berkembang menjadi ilmu pengetahuan, di mana akarnya ditanam dengan kokoh dan cabang-cabangnya menjulang tinggi, sehingga mampu menghasilkan hasil yang terbaik dengan izin Tuhan. Menghafal, selain sebagai cara untuk memperoleh pemahaman tentang suatu materi pelajaran, juga berfungsi sebagai cara untuk meningkatkan budaya membaca di kalangan siswa.

Minat dan bakat siswa memiliki perbedaan yang nampak jelas dari keseriusannya dalam mempelajari mata pelajaran yang sedang dia pelajari, terkadang beberapa siswa hanya fokus pada mata pelajaran kesukaannya, dan tak jarang ketika mata pelajaran yang tidak dia sukai sedang diajarkan, dia enggan memperhatikan atau malah takut dan kabur untuk masuk ke proses pembelajaran tersebut, hal ini membuat prestasi siswa menjadi tidak merata dan membuat hasil dari proses pendidikan selama satu semester menjadi tidak efisien, sehingga perlu melakukan remedial untuk memperbaiki nilai mata pelajaran yang

memiliki nilai rendah, hal ini membuat guru sedikit kewalahan dan menjadi masalah setiap kali semester akan berakhir, berangkat dari masalah yang ada, penelitian ini berharap agar kondisi ini bisa diatasi dengan pendeteksian dini kepada siswa untuk mengetahui siswa mana yang perlu perhatian lebih, agar kedepannya setiap siswa dapat mencapai target hafalan Al-Quran.

Penelitian mengenai faktor dalam menghafal telah banyak dilakukan, seperti penelitian yang mengukur tingkat hafalan berdasarkan faktor behavior untuk meningkatkan motivasi siswa dalam menghafal. (Marlina, 2018) pada penelitian ini telah didapatkan fakta bahwa pemberian kesadaran yang memotivasi siswa untuk menghafal, dapat meningkatkan tingkat ketekunan siswa dalam menghafal, sehingga siswa dengan motivasi tinggi lebih cepat prosesnya dalam meraih target hafalan dibanding siswa yang motivasinya rendah atau tidak termotivasi sedikitpun.

Pada proses menghafal telah diukur faktor yang mempengaruhinya seperti pengaruh komunikasi interpersonal guru dan siswa dalam upaya peningkatan hafalan dalam bentuk motivasi (Desti dan Yunidar, 2021) (Oktarina and Yanti, 2021) interaksi guru dan murid melalui sebuah komunikasi terbukti mempengaruhi tingkat motivasi siswa dalam menghafal, berdasarkan pengukuran yang telah dilakukan dapat diungkapkan bahwa 33,3% dari variabel komunikasi interpersonal memiliki pengaruh pada perhatian terhadap hafalan ayat suci Al-Qur'an.

Dalam menghafal ayat suci Al-Qur'an, faktor eksternal serta faktor internal dapat mempengaruhi seberapa mudah atau sulitnya proses hafalan itu

sendiri. Faktor kepribadian sebagai pembeda dari setiap individu menjadi faktor internal yang sangat berpengaruh pada proses menghafal ayat suci Al-Qur'an, kemudian teknik yang diterapkan dalam proses hafalan merupakan faktor dari luar yang akan berpengaruh sama pada setiap individu yang menggunakannya (Fitriyah, 2008)

Faktor minat siswa dalam menghafal berpotensi besar dalam kelancaran proses penghafalan Al-Qur'an (Abbas dan M. Ziyad, 1993) (Alfatoni dan Sabit, 2019), kemudian faktor motivasi dan stimulus (perubahan yang dapat diketahui berupa faktor lingkungan internal atau eksternal) menjadi faktor yang mempengaruhi proses hafalan siswa (Abdurrah Nawabuddin dan Bambang Saiful Ma'arif, 2005).

Selain itu faktor usia, lokasi, dan manajemen waktu menjadi faktor yang cukup berpengaruh dalam menghafal (Ahsin Al-Hafidz, 2005). Selain itu, jumlah kata dan panjang ayat dari Al-Qur'an mempengaruhi waktu yang diperlukan untuk menghafalnya (Abdul Khaliq dan Syaikh Abdurrahman, 2013).

Berdasarkan penelitian sebelumnya telah diketahui bahwa faktor internal yang mempengaruhi hafalan siswa adalah kepribadian, minat dan bakat, serta kemampuan daya ingat (memori), sedangkan untuk faktor eksternal diantaranya metode yang digunakan, motivasi, kelancaran membaca dan menulis, amalan yang rutin dilakukan, waktu, fasilitas dan tempat. Berdasarkan hal ini maka peneliti akan menjadikan faktor tersebut sebagai variabel acuan dalam mengukur tingkat hafalan siswa.

Faktor-faktor seperti motivasi dan rangsangan (perubahan lingkungan

dalam atau luar yang dapat diidentifikasi) juga memainkan peran penting dalam proses menghafal siswa. Usia, lokasi, dan manajemen waktu juga merupakan faktor yang cukup berpengaruh dalam proses menghafal (Wijaya et al., 2021)

Kemudian penelitian yang dilakukan untuk mengukur prediksi pencapaian sudah banyak diteliti sebelumnya, seperti melakukan pengujian untuk mendapatkan faktor apa saja yang paling berpengaruh pada pencapaian atau kelulusan siswa (Christin Nandari Dengen, dkk, 2019). Sebagian besar dari data siswa berupa nilai menjadi penyebab terpenting dalam menentukan apakah siswa tersebut mampu menyelesaikan atau mencapai tujuan yang diharapkan, sedangkan faktor seperti jenis kelamin, tempat tinggal tidak begitu mempengaruhi hasil akhir dari prediksi kelulusan siswa. Berdasarkan kumpulan data yang ada maka perlu untuk dilakukan pemanfaatan guna mendapatkan informasi mengenai siswa mana yang dapat mencapai target dalam menghafal Al-Quran, Untuk mencari tahu hal ini diperlukan rancangan model data mining yang dapat memprediksi siswa mana yang dapat mencapai target hafalan atau tidak

Data mining merupakan sebuah proses dengan tujuan mendapatkan informasi penting atau pola menarik yang sumbernya adalah sejumlah data yang sangat besar yang disebut big data (Sugiharti, 2017).

Proses data mining terdiri dari beberapa langkah, termasuk tahap pra-pemrosesan, pemrosesan, dan pasca-pemrosesan. Dalam tahap pra-pemrosesan terdapat beberapa sub-langkah, seperti membersihkan data, mengintegrasikan data, mengurangi ukuran data, dan mengubah bentuk data (Mabrur dan Lubis,

2012).

Meskipun sejumlah besar data yang kompleks dapat memberikan informasi yang berguna, jumlah atribut pada data tersebut dapat menurunkan keakuratan dan mempersulit langkah dalam menggunakan algoritma untuk mengolah data tersebut. Dalam proses pemilihan atribut untuk mendapatkan hal apa saja yang paling mempengaruhi dalam proses perhitungan pencapaian siswa dibutuhkan sebuah langkah untuk menyeleksi data (Elin dan Abdul, 2019). Kumpulan data bisa menghasilkan informasi baru, namun di sisi lain apabila kita tidak menggunakan data yang tepat dalam menentukan prediksi pencapaian siswa, hal itu hanya akan memperlambat proses perhitungan.

Penelitian mengenai fitur seleksi ini telah banyak dilakukan salah satu manfaat yang didapatkan ialah didapatkan data yang lebih mudah untuk dilakukan perhitungan dan meningkatkan akurasi (I Made Budi Adnyana, 2019). Dalam proses perhitungan machine learning fitur seleksi data menjadi sangat penting untuk menghadapi sebuah data yang kompleks.

Pemilihan fitur diperlukan untuk meningkatkan kinerja machine learning dalam mengukur tingkat keberhasilan hafalan ayat suci Al-Qur'an siswa. Terdapat beberapa teknik untuk memilih fitur, hal ini dikemukakan oleh para peneliti, seperti metode backward feature elimination dan forward feature selection (Meyer, dkk, 2018)(Balram, dkk, 2019)(Bergman, dkk, 2018). Dalam penelitian ini, akan mengaplikasikan forward feature selection untuk memperbaiki akurasi prediksi pada machine learning.

Pada penelitian ini nantinya akan diterapkan Algoritma Forward Selection, hal ini berdasarkan studi literatur dan penjelasan dari berbagai sumber yang menyimpulkan bahwa Forward Selection mendapatkan akurasi yang lebih baik dengan cara memasukkan satu persatu atribut ke dalam model, dibandingkan Backward Selection yang dimana semua atribut dimasukkan kemudian setiap atribut diseleksi satu persatu, hal ini terbukti tidak lebih baik berdasarkan penelitian terdahulu.

Untuk menyelesaikan masalah optimisasi, algoritma Particle Swarm Optimization (PSO) merupakan salah satu metode heuristik yang dapat mengatasi permasalahan tersebut (Ashari, 2016: 149). Pada beberapa pengujian, terbukti bahwa PSO sangat kompetitif dibandingkan dengan algoritma lainnya (Muslim, 2017: 2). Metode optimisasi ini terbukti efektif untuk mengatasi masalah optimisasi yang memiliki beberapa dimensi dan parameter dalam bidang pendidikan machine learning, seperti jaringan saraf tiruan dan teknik-teknik klasifikasi algoritma (Fei, 2009: 1605).

K-Nearest Neighbor (KNN) adalah sebuah teknik yang mengklasifikasikan objek baru berdasarkan jarak terdekat ke data pembelajaran yang ada. Ini adalah metode yang sederhana dan mudah diimplementasikan, seperti teknik pengelompokan yang mengklasifikasikan data baru berdasarkan jarak terdekat ke tetangga-tetangga. Jumlah tetangga yang digunakan dalam proses klasifikasi harus ditentukan sebelum mencari jarak data ke tetangga (Dzikrullah, 2017: 380).

Berdasarkan studi literatur yang dilakukan sebelumnya, pengujian ini

bertujuan mengukur serta mengembangkan model untuk memprediksi pencapaian target hafalan ayat suci Al-Qur'an siswa. Dengan pertimbangan nilai kelas mata pelajaran umum secara keseluruhan kemudian digabungkan, setelah itu dikonversi menjadi satu angka, selain itu faktor-faktor yang berkaitan dalam mempengaruhi proses hafalan ayat suci Al-Qur'an akan diidentifikasi pada penelitian ini. seperti faktor internal diantaranya nilai kepribadian atau sikap, lalu untuk faktor eksternal seperti nilai kemampuan beradaptasi siswa dengan metode yang digunakan, motivasi(dilihat dari seberapa sering siswa menyetorkan hafalan), kelancaran membaca, menulis, dan tingkat kerajinan mengerjakan amalan-amalan lainnya seperti sholat sunnah, dan dzikir, Algoritma yang diaplikasikan pada penelitian ini diantaranya algoritma K-Nearest Neighbor (KNN), algoritma Particle Swarm Optimization (PSO) yang berfungsi yang untuk menghitung bobot dari setiap variabel untuk mendapatkan solusi atribut optimal, dan untuk tahap seleksi data guna mengurangi variabel yang tidak begitu berpengaruh besar atau redundant, maka diterapkan Forward Selection (FS).

Berangkat dari studi literatur, maka penelitian ini dilakukan untuk mengukur dan mendapatkan model untuk memprediksi target hafalan siswa. Dengan mengolah faktor-faktor yang mempengaruhi proses siswa dalam menghafalkan ayat suci Al-Qur'an. Algoritma yang digunakan dalam penelitian ini adalah algoritma K-Nearest Neighbor (KNN), algoritma Particle Swarm Optimization (PSO) yang berfungsi yang untuk menghitung bobot dari setiap variabel untuk mendapatkan solusi atribut optimal, dan untuk tahap seleksi data

guna mengurangi variabel yang tidak begitu berpengaruh besar atau redundant, maka diterapkan algoritma Forward Selection (FS). Penelitian ini berjudul “Particle Swarm Optimization Untuk Penentuan Parameter dan Fitur Seleksi Pada Algoritma K-Nearest Neighbor Pada Prediksi Target Hafalan Siswa SMPIT MIS”

### **1.2. Rumusan Masalah**

Berdasarkan latar belakang masalah yang telah dijelaskan, maka rumusan masalah dalam penelitian ini sebagai berikut:

- a. Berapa tingkat akurasi algoritma KNN dalam memprediksi target hafalan siswa?
- b. Berapa tingkat akurasi dan efisiensi pada akurasi dari algoritma KNN setelah dioptimasi dengan PSO dan FS dalam memprediksi target hafalan siswa?
- c. Variabel apa yang paling dominan mempengaruhi akurasi dalam perhitungan target hafalan siswa baik sebelum maupun sesudah diterapkannya algoritma PSO dan FS?

### **1.3. Batasan Masalah**

Pembatasan masalah diterapkan agar menghindarkan dari kemungkinan adanya penyimpangan ataupun meluasnya inti dari pokok masalah dan diharapkan agar penelitian lebih terarah serta memudahkan dalam pembahasan sehingga tujuan penelitian dapat tercapai. Beberapa batasan masalah dalam penelitian ini adalah sebagai berikut:

- a. Data yang dipakai pada penelitian ini hanya data siswa pada SMP MIS.
- b. Penelitian ini hanya akan memperhatikan variabel yang telah ditentukan.
- c. Penelitian ini tidak memperhatikan selain nilai berupa angka.
- d. Penelitian ini tidak memperhatikan prestasi siswa di luar sekolah.
- e. Penelitian ini hanya melakukan optimasi dengan PSO dan Forward Selection dalam meningkatkan akurasi algoritma KNN
- f. Penentuan jumlah *dataset* yaitu dengan pengambilan data sampel dengan mengacu pada nilai siswa (persemester).
- g. Penentuan keputusan akhir mengacu pada hasil yang dilakukan oleh KNN.
- h. Pengukuran perbandingan akurasi menggunakan *confussion matrix*

#### **1.4. Tujuan Penelitian**

Secara umum, tujuan penelitian ini adalah sebagai berikut:

- a. Mengetahui performa akurasi algoritma KNN dalam memprediksi apakah siswa mampu mencapai target hafalan Al-Qur'an atau tidak.
- b. Membandingkan hasil perhitungan pada proses prediksi pencapaian target hafalan Al-Qur'an siswa dengan algoritma KNN, sebelum dan sesudah diterapkan algoritma Particle Swarm Optimization (PSO) dan fitur Forward Selection pada data siswa.
- c. Menentukan variabel apa yang paling mempengaruhi akurasi.

### 1.5. Manfaat Penelitian

Bagian ini memuat penjelasan tentang:

- a. Bagi Sekolah (SMP MIS), dapat mengelola pembelajaran peserta didik agar lebih efisien dan terarah serta mendapatkan peningkatan prestasi secara signifikan, dan bisa memberikan pengarahannya berupa konseling ke siswa yang dianggap perlu untuk mendapat perhatian lebih.
- b. Bagi Siswa (SMP MIS), dapat mengetahui kekurangan dirinya pada pencapaiannya dalam menghafal Al-Quran sehingga bisa melakukan manajemen waktu agar hafalan mereka tetap bisa mencapai target
- c. Bagi peneliti dapat mengetahui manfaat dari penerapan PSO dan Forward Selection pada Algoritma KNN untuk keperluan pendidikan, khususnya pada prestasi peserta didik, serta mendapatkan pemahaman bahwa optimasi menggunakan PSO dan Forward Selection mampu diterapkan pada algoritma KNN, serta mengetahui performa dari algoritma KNN setelah diterapkan PSO dan Fitur Forward Selection.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Tinjauan Pustaka

Ada beberapa riset yang telah dilakukan untuk membahas prediksi kelulusan atau pencapaian siswa. salah satunya ialah penelitian mengenai prediksi kelulusan mahasiswa yang dimana pada penelitian ini akan dibuat sebuah proses data mining untuk memprediksi kemungkinan mahasiswa untuk lulus tepat waktu atau tidak (Dengen, dkk, 2019). Salah satu keunggulan dari penelitian ini adalah penggunaan model standarisasi data mining yang dikenal sebagai Cross Industry Standard Process for Data Mining (CRISP-DM) untuk memetakan langkah pembuatan model data mining kemudian akan menggunakan algoritma Decision Tree (DT) sehingga alurnya menjadi jelas dan keberhasilan penerapan algoritma mampu memberikan informasi baru dalam menentukan kapan mahasiswa dinyatakan atau diprediksi mampu untuk mencapai kelulusan tepat waktu, atau terlambat. Namun kelemahan pada penelitian ini ialah kurangnya populasi pada *dataset* sehingga data yang dikelola kurang beragam sehingga model yang diterapkan tidak bekerja secara maksimal. Penelitian ini bertujuan untuk menggunakan metode decision tree untuk memprediksi keberhasilan mahasiswa dalam menyelesaikan studi sesuai waktunya di Universitas Mulawarman pada Program Studi Teknik Informatika, dengan tingkat akurasi 60%. Selain itu, hasil penelitian dapat digunakan sebagai panduan oleh program studi untuk mengambil tindakan yang tepat dalam meningkatkan keberhasilan mahasiswa dalam menyelesaikan studi tepat waktu.

Penelitian lain terkait upaya dalam meminimalisir kegagalan dalam studi ialah dengan melakukan prediksi sedini mungkin sehingga dapat mengetahui apa faktor penghambat dalam kelulusan mahasiswa (Endang, Dkk. 2020). Kelebihan dari penelitian ini ialah faktor-faktor yang menghambat dideteksi dengan cara mengumpulkan data mahasiswa kemudian melakukan klasifikasi terhadap data tersebut yang menghasilkan hasil prediksi berupa pernyataan apakah mahasiswa mampu mencapai target kelulusan tepat waktu atau tidak, hal lain yang menjadi kelebihan dari penelitian ini ialah penulis melakukan perbandingan dengan banyak algoritma seperti Naive Bayes (NB) dan algoritma C4.5. Kelemahan dari penelitian ini ialah jumlah atribut yang dinilai cukup sedikit serta penambahan faktor eksternal, jadi data yang dimiliki menjadi lebih beragam, sehingga faktor internal dan eksternal menjadi variabel atau kriteria yang dapat dinilai secara bersamaan dalam menentukan tingkat keberhasilan mahasiswa, misalnya status pekerjaan, dengan cara apa mereka membayar uang kuliah, apakah mereka sudah menikah atau belum. Kesimpulan dari penelitian ini menunjukkan bahwa metode algoritma C4.5 lebih efektif dalam memprediksi kelulusan mahasiswa di STMIK Bina Nusantara Jaya Lubuklinggau, dengan nilai akurasi sebesar 79,08%. Selain itu, pohon keputusan yang diimplementasikan menunjukkan bahwa variabel IPK-S4, jenis kelamin dan asal sekolah mempengaruhi prediksi kelulusan. Disarankan untuk menggunakan metode ini untuk menyelesaikan masalah prediksi kelulusan siswa di STMIK Bina Nusantara Jaya Lubuklinggau.

Terdapat beberapa penelitian lain yang telah dilakukan untuk mengetahui

kapan siswa akan lulus tepat waktu (Budiyanto dan Fatimah, 2019). Penelitian ini mengungkapkan bahwa tingkat kelulusan tepat waktu mahasiswa merupakan informasi penting bagi institusi, karena dapat memberikan manfaat seperti peningkatan kualitas institusi dan nilai akreditasi. Oleh karena itu, memprediksi ketepatan waktu kelulusan mahasiswa dengan cepat akan sangat bermanfaat bagi institusi serta mahasiswa. Pada penelitian Jaringan Syaraf Tiruan digunakan untuk memprediksi ketepatan waktu kelulusan mahasiswa, dengan dataset yang cukup banyak. Namun, ada beberapa kelemahan dalam proses pengolahan data yang tidak mengalami peningkatan, sehingga hasil yang dihasilkan kadang tidak begitu tepat dalam hal akurasi prediksi kelulusan. Namun, setelah diuji menggunakan confusion matrix, nilai akurasi yang diperoleh adalah 0,805921053, yang menunjukkan bahwa penelitian ini cukup baik dan mampu melakukan prediksi.

Dalam memprediksi kelulusan siswa. Oleh karena itu, dibutuhkan metode untuk menyaring atribut tidak relevan dari atribut yang penting dalam proses prediksi kelulusan siswa tepat waktu. Data mining menyediakan metode seleksi fitur yang dapat digunakan untuk menyaring atribut yang tidak penting dan meningkatkan akurasi dalam proses prediksi kelulusan siswa tepat waktu. (Christin, dkk, 2019). Ada banyak hal yang perlu diketahui atau dipertimbangkan dalam menentukan atribut dalam pengumpulan data, sehingga ketika sebuah data digunakan pada proses data mining, hal itu bisa meningkatkan akurasi pada perhitungan, untuk mengetahui faktor-faktor yang mempengaruhi kelulusan siswa tepat waktu, diperlukan penelitian yang difokuskan pada identifikasi

faktor-faktor yang memiliki pengaruh penting dalam perhitungan. Hal ini dapat dilakukan dengan menentukan asosiasi antara faktor-faktor terkait dengan data yang dikelola, penelitian tersebut diharapkan bisa meminimalisir penggunaan atribut yang tidak dibutuhkan pada proses prediksi kelulusan atau pencapaian siswa.

Pada sebuah data, sering ditemukan data yang tidak seimbang antara data yang positif dan data negatif atau dalam hal ini antara data Untuk mengetahui perbedaan antara siswa yang berhasil mencapai target dan siswa yang tidak, diperlukan penelitian untuk mengidentifikasi faktor-faktor yang mempengaruhi kinerja siswa dan membedakan antara kedua kelompok tersebut. (Achmad dan Rinna, 2019). Maka pada data yang seperti ini juga perlu untuk diolah agar hasil penilaian dapat lebih baik, Hal lain yang menjadi masalah ketika data yang ingin dikelola terlalu sedikit sehingga ketika pengujian data menggunakan model atau algoritma maka dapat terjadi *overfitting*, maka data perlu untuk divalidasi (Lalu, dkk, 2019). Proses validasi data merupakan langkah untuk mendefinisikan *dataset* yang berfungsi pada pengujian model dalam fase pelatihan, Untuk menghindari masalah seperti *overfitting*, tindakan yang dapat diambil adalah melakukan validasi silang untuk memberikan pandangan tentang bagaimana model akan digeneralisasikan secara independen dari dataset yang digunakan. Jumlah data dengan atribut yang tidak dibutuhkan dapat membuat proses perhitungan menjadi rumit dan menghambat kinerja dari sebuah algoritma (Saputra dan Abdul, 2019) Melakukan seleksi data dan menerapkan optimasi pada algoritma agar mendapatkan hasil yang optimal bisa menjadi langkah yang

baik untuk meningkatkan akurasi . dalam penelitian yang mereka lakukan terlabat kelebihan yaitu dapat diketahui secara detail bahwa penerapan optimasi PSO mampu menyeleksi data sehingga data yang diolah menjadi lebih baik dan akurasi mengalami peningkatan, namun kekurangan dari penelitian ini adalah *dataset* yang digunakan kurang beragam sehingga pemanfaatan dari PSO belum bisa maksimal. Melalui pengujian yang dilakukan dengan mengaplikasikan algoritma Neural Network (NN), didapatkan akurasi 94.00% serta diperoleh AUC senilai 0.914. Namun, pada mengaplikasikan metode NN yang didukung oleh PSO untuk seleksi atribut dan pengaturan parameter dan populasi, didapatkan hasil yang lebih baik. Dua atribut yang terpilih dari pengujian ini adalah latihan dan chat. Nilai akurasi yang diperoleh mencapai 98.50% serta didapatkan AUC senilai 0.997. Hal ini membuktikan bahwa metode ini lebih efektif dalam menentukan tingkat kelulusan mahasiswa berdasarkan atribut yang dipilih. Kategori klasifikasi akurasi yang didapatkan masuk pada kategori "sangat baik" yaitu nilai akurasi meningkat sebesar 4,5% serta AUC senilai 0.83.

Berdasarkan penelitian sebelumnya terkait dengan prediksi pencapaian mahasiswa apakah mampu untuk lulus tepat waktu atautkah tidak, sehingga peneliti termotivasi untuk melakukan pengujian selanjutnya terkait dengan ide mengenai prediksi apakah siswa mampu menghafal Al-Quran sesuai target yang ditetapkan yang ditinjau dari berbagai aspek di SMPIT MIS, kemudian pada proses pengolahan data, maka akan diterapkan proses data mining, disini penulis menggunakan metode yang berbeda namun juga berdasarkan kajian dari peneliti sebelumnya tentang penggunaan metode tersebut. Berikut beberapa kajian

tentang penggunaan metode tersebut mencakup dengan seleksi data seperti FS, algoritma optimasi seperti PSO dan algoritma KNN.

Berdasarkan penelitian yang dilakukan oleh peneliti sebelumnya dalam tahap seleksi data menggunakan FS maka dari itu penulis meninjau hasil penelitian sebelumnya yang menggunakan FS dengan tujuan untuk menerapkannya kembali. Tujuan tahap pemilihan data untuk mengurangi jumlah data yang akan dipakai dalam proses perhitungan dengan tetap mewakili data asli. FS digunakan pada tahap seleksi data sebelum data digunakan untuk tahap selanjutnya. Forward Selection digunakan untuk menyelesaikan masalah data yang tidak relevan dengan menghilangkan atribut yang tidak penting dari proses mining data (J. Han dan M. Kamber, 2006). Algoritma FS berdasarkan pada suatu metode yang menggunakan model regresi linear untuk mengevaluasi kontribusi setiap atribut dalam menentukan hasil yang diinginkan. Dengan demikian, algoritma ini diaplikasikan untuk mengurangi atribut-atribut yang dianggap tidak relevan dari data yang digunakan dalam proses mining (R. Noori, dkk, 2011). Algoritma ini diterapkan untuk mereduksi besaran *dataset* dengan mengurangi atribut yang dianggap tidak relevan atau redundan (Hermawati dan Nuruddin, 2016). Penelitian yang dilakukan oleh Hermawati mengevaluasi sebuah model yang dapat mendeteksi diabetes dengan menggunakan dataset diabetes di laboratorium dan menggunakan algoritma KNN setelah data yang tidak relevan dihilangkan. Keuntungan dari penelitian ini adalah data yang digunakan telah dibersihkan, namun kelemahannya adalah algoritma KNN tidak dioptimalkan. Menggabungkan Forward Selection dan KNN dalam algoritma

meningkatkan keakuratan dalam mendeteksi diabetes, dengan tingkat keberhasilan sebesar 96,08%.

Prediksi kelulusan siswa dengan jumlah data yang banyak menjadi hal yang sulit untuk dilakukan oleh manusia, terdapat data pada *database* milik institusi yang dalam hal ini bisa kita manfaatkan untuk memprediksi siswa mana yang mampu lulus tepat waktu dengan penerapan *data mining*, berdasarkan penelitian terkait yang telah diteliti metode KNN merupakan algoritma terbaik yang dapat diterapkan untuk memprediksi kelulusan (Imaning, dkk, 2019). Selain digunakan dalam diagnosa penyakit diabetes, metode KNN juga terbukti efektif dalam proses klasifikasi pemilihan bibit sapi Bali yang unggul (Putra dkk, 2015) serta pemilihan makanan yang bergizi (Afandie, 2014). Penelitian yang dilakukan oleh imaning meneliti pengaruh nilai K pada algoritma KNN terhadap tingkat akurasi. Penelitian ini menggunakan algoritma KNN yang telah dimodifikasi untuk meningkatkan akurasi. Hasil dari pengujian ini memperoleh fakta bahwa nilai K terbaik untuk mencapai akurasi tertinggi adalah 5, dengan tingkat akurasi sebesar 82%. Penelitian ini juga menunjukkan bahwa jumlah data-latih yang digunakan dan komposisi data latih juga mempengaruhi tingkat akurasi, dengan akurasi tertinggi sebesar 85,24% ketika menggunakan 140 data latih dan 84% ketika menggunakan komposisi data 50 kelas TW dan 50 kelas TTW. Kelebihan dari penelitian ini adalah algoritma KNN yang telah dimodifikasi menghasilkan akurasi yang lebih tinggi, sementara kekurangannya adalah ketidakseimbangan data yang tidak ditangani oleh peneliti. Dapat disimpulkan bahwa penggunaan algoritma Modified MK-NN mampu memperoleh akurasi lebih optimal

dibandingkan penerapan KNN saja.

Penerapan algoritma KNN telah banyak diuji oleh penguji sebelumnya seperti penelitian terkait prediksi kelulusan menggunakan data registrasi mahasiswa (Nabila, dkk, 2021). Penelitian yang dilakukan oleh Nabila berfokus pada perbaikan proses registrasi mahasiswa baru dengan menggunakan model prediksi yang dibuat memakai metode Fuzzy C-Means dan KNN. Dalam pembuatan model ini, data dikelompokkan dan diberi label dengan menggunakan FCM, kemudian jarak antar data ditentukan dengan menggunakan KNN, yang membuat proses jarak  $k$ -tetangga terdekat hanya dilakukan pada data yang berada dalam satu cluster saja, sehingga model tidak menjadi rumit. Kelebihan dari penelitian ini adalah data dikelompokkan terlebih dahulu sebelum dilakukan perhitungan, sehingga informasi yang ada dikelola sesuai dengan kelompoknya. Namun, penelitian ini tidak melakukan perbandingan dengan algoritma lain yang mungkin dapat menjadi penunjang dalam penerapan model yang dirancang. Hasil dari penerapan algoritma FCM-KNN menunjukkan bahwa model memiliki rata-rata akurasi sebesar 71% saat menerapkan pengujian 10-fold cross validation pada skenario  $k$  bernilai 1, dan nilai akurasi akan berubah ketika nilai  $K$  (tetangga terdekat) bertambah.

Penelitian lain mengenai perbandingan algoritma Naive Bayes dan KNN yang dimana kasus yang dihadapi ialah prediksi bagaimana kriteria mahasiswa yang dapat lulus tepat waktu atau mengalami keterlambatan (Sidik Rahmatullah dan Ema Utami, 2019). Pada penelitian ini terdapat kelebihan di mana kasus dihadapkan atau dicarikan solusi dengan dua model, untuk menemukan jawaban

yang paling cocok untuk mengatasi masalah yang dihadapi, namun terdapat kelemahan pada penelitian ini yang di mana algoritma tidak mengalami optimasi, sedangkan pada tahap persiapan data sama sekali tidak mengalami perbaikan data, yang kemungkinan data yang diolah bisa saja diterapkan *cleaning, selection* dan hal serupa yang mungkin bisa memperbaiki kondisi data sebelum digunakan, kemudian pada tahap pemisahan data latih dan data uji diterapkan secara manual, yang dimana hal ini menjadikan data tidak dimanfaatkan sepenuhnya, sebab terdapat data yang tidak mengalami pelatihan dengan model, begitupun sebaliknya, terdapat data yang tidak digunakan untuk pengujian model. Hasil dari pengujian menunjukkan fakta bahwa dengan penerapan Naive Bayes menghasilkan nilai tingkat akurasi sebesar 85%, sedangkan ketika menggunakan algoritma KNN akurasi didapatkan 68.89 %.

Penelitian lain yang mencoba membandingkan beberapa algoritma untuk menangani kasus prediksi kelulusan mahasiswa (Sri Widaningsih, 2019). Keunggulan dari penelitian ini adalah digunakannya beberapa algoritma dalam pengujiannya, yaitu C4.5, SVM, kNN, dan Naive Bayes. Proses pengolahan data awal dilakukan melalui KDD yang meliputi beberapa tahap seperti pemilihan, pra-pemrosesan, transformasi, penambahan data dan interpretasi/evaluasi. Namun, kelemahan dari kajian ini adalah kurangnya faktor yang mempengaruhi hasil prediksi. Hasil akhir dari uji coba menunjukkan bahwa algoritma Naive Bayes adalah yang terbaik dalam menentukan mahasiswa mampu lulus tepat waktu dan memiliki IPK sama dan lebih besar dari 3, dengan tingkat keakuratan sebesar 76,79%, kesalahan sebesar 23,17% dan AUC sebesar 0,850.

Ketika meninjau dari penelitian sebelumnya maka algoritma KNN kurang begitu bekerja dengan baik sebab hasil akurasi tidak lebih unggul dibanding algoritma lain, sehingga penting untuk melakukan optimasi pada algoritma ini, serta penentuan nilai K menjadi sangat penting yang di mana nilai tersebut sangat mempengaruhi nilai akurasi yang ada, Terdapat penelitian yang menerapkan fitur optimasi yaitu PSO yang berfungsi untuk menyeleksi atribut sesuai dengan kebutuhan pengolahan data nantinya (Elin Panca Saputra dan Abdul Hamid, 2019). Kelebihan dari penelitian ini, peneliti mencoba menyeleksi data atribut menggunakan PSO, kemudian mengolah data menggunakan Neural Network (NN) kelebihan dari NN ialah mampu menghitung nilai dengan baik meskipun dengan nilai sampel yang sedikit, sehingga penerapan PSO akan mampu mengurangi data berlebih yang bisa membebani kinerja dari algoritma NN. Kekurangan dari penelitian ini adalah kurangnya variasi dalam dataset yang digunakan, sehingga penerapan algoritma PSO belum dapat dilihat hasilnya secara maksimal. Hasil yang diperoleh dari penelitian pada penerapan algoritma NN menunjukkan tingkat keakuratan yang tinggi senilai 94% dengan nilai AUC 0,914%. Namun, dengan mengaplikasikan metode algoritma Neural Network Berbasis PSO dan pemilihan atribut dari 7 variabel prediktor, diperoleh hasil yang lebih baik dengan tingkat keakuratan sebesar 98,50% dan AUC senilai 0,997%. Ini menandakan bahwa dengan mengaplikasikan algoritma Neural Network berbasis PSO, dapat menemukan kategori klasifikasi yang lebih tinggi dengan pemilihan atribut dari 7 variabel prediktor yang diuji dan atribut latihan serta chat merupakan yang terpilih.

Pada penelitian lain terdapat penerapan kelulusan mahasiswa menerapkan algoritma Support Vector Machine dengan optimasi PSO yang dimana berfungsi untuk mengatasi optimasi pada parameter yang terdapat pada data yang diolah (Suhardjono, dkk, 2019). Dalam penelitian yang dilakukan oleh Suhardjono, ditemukan bahwa menggunakan metode Particle Swarm Optimization dapat mengatasi masalah yang muncul dalam proses optimisasi parameter pada Support Vector Machine. Kelebihan penelitian ini adalah terdapat pada optimasi yang dilakukan, yang di mana dinilai mampu memperbaiki perhitungan yang dilakukan. Kekurangan pada penelitian ini ialah terdapat atribut IPK yang tidak memiliki pengaruh terhadap perhitungan seperti IPK pada semester pertama dan ke empat yang dimana pada perhitungan tidak memiliki bobot yang berpengaruh terhadap peningkatan akurasi. Hasil yang diperoleh dalam pengujian ini membuktikan bahwa PSO mampu memperbaiki kinerja Support Vector Machine (SVM) dalam mengoptimalkan parameter sebesar 00.62%, atau yang awalnya hanya sebesar 85.81% menjadi 86.43% sehingga dalam kasus yang diselesaikan yaitu prediksi kelulusan mahasiswa, hasil akurasi ini dapat berpengaruh besar.

Pada kasus yang sama dengan penelitian sebelumnya namun menggunakan algoritma Decision Tree (DT) Berbasis PSO, pada pengujian ini PSO berfungsi menghilangkan fitur yang tidak digunakan (Hendra, dkk, 2020). Kelebihan penelitian ini adalah dapat dilihat pada peran Particle Swarm Optimization dalam mengatasi kelemahan yang ada pada algoritma Decision Tree yang di mana PSO mampu menghilangkan parameter tidak terlalu penting

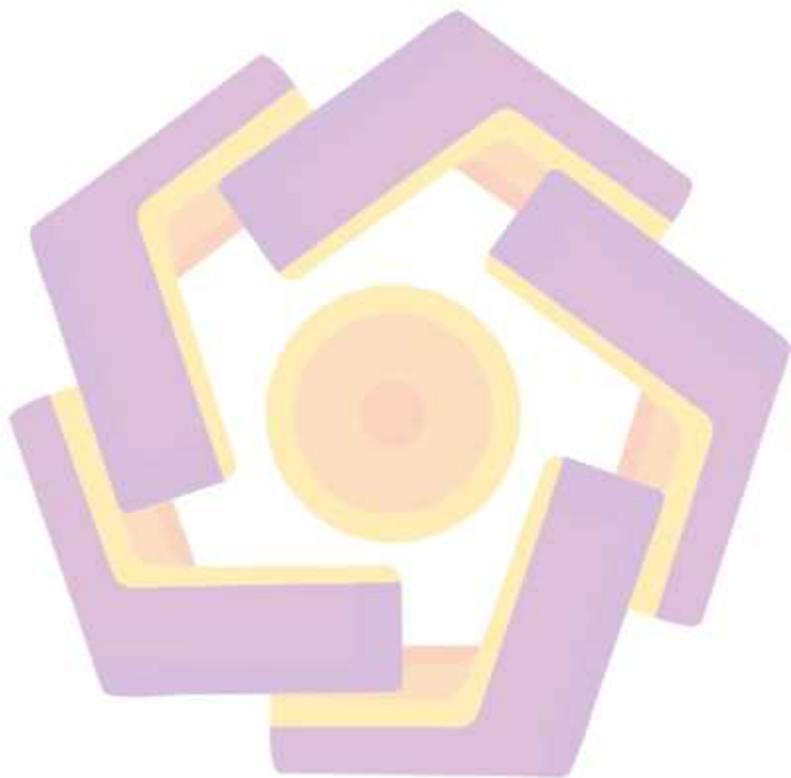
yang di mana menghasilkan satu set data yang efektif. Kelemahan penelitian ini adalah masih sangat banyak prediksi yang salah, terdapat 99 data mengalami kegagalan dalam prediksi dari 796 data yang digunakan. Hasil penelitian yang dilakukan di Amik PPMI Tangerang membuktikan penerapan metode PSO pada data primer mahasiswa yang lulus antara tahun 2000 hingga 2003 dapat meningkatkan tingkat keakuratan sebesar 87,56%, yang merupakan peningkatan sebesar 1,01% dibandingkan dengan metode Decision Tree yang memiliki tingkat keakuratan sebesar 86,55%. Selain itu, metode ini memungkinkan bisa mengetahui atribut mana yang tidak digunakan dan tidak memiliki bobot yang signifikan, sehingga dapat mempengaruhi hasil akhir. Penemuan ini dapat digunakan untuk mempercepat proses kelulusan siswa di Amik PPMI Tangerang.

Berdasarkan penelitian sebelumnya didapatkan fakta bahwa ketika KNN diterapkan, maka sangat penting untuk menerapkan optimasi dan perbaikan pada tahap pengolahan data, sebab akan sangat mempengaruhi hasil akurasi dari model yang dibuat, algoritma ini ketika dibandingkan dengan algoritma lainnya tanpa mengalami optimasi, maka nilai akurasi yang dihasilkan cukup rendah sehingga performa dari rancangan model menjadi kurang baik dan kurang efektif untuk mengatasi masalah yang ingin dipecahkan, kemudian berdasarkan penelitian yang ada mengenai penerapan algoritma Particle Swarm Optimization, didapatkan fakta bahwa optimasi ini mampu untuk mengatasi perbaikan pada bobot *dataset* yang digunakan sehingga kita mampu mengetahui atribut mana saja yang mampu memberi kontribusi pada peningkatan akurasi dari

penerapan pengolahan data menggunakan teknik data mining, atau dalam hal ini prediksi pencapaian target siswa yang ditinjau dari pencapaian nilai-nilai selama belajar.

Prediksi pencapaian siswa menjadi sangat penting, sebab hal ini menjadi langkah yang diambil oleh institusi untuk memperbaiki sistem pendidikan yang ada sesegera mungkin, semisal dengan bimbingan konseling, nasihat, pembelajaran tambahan, perhatian khusus dan sebagainya, sementara itu untuk mewujudkannya dibutuhkan prediksi untuk mengetahui siswa mana saja yang mampu mencapai target, dan siswa mana saja yang mengalami hambatan dalam mencapai target yang telah ditetapkan. Ada banyak faktor yang dapat menghambat siswa dalam mencapai target pembelajaran, misalnya penggunaan *smartphone* secara berlebihan, bermain *game*. Dan kurang fokus pada pembelajaran, hal itu bisa berdampak baik untuk kondisi psikologis siswa, namun bisa berdampak buruk pada pendidikan ketika hal itu tidak terkontrol dengan baik. Berdasarkan penelitian yang ada dapat dilihat bahwa nilai optimasi PSO, Hasil dari penelitian mengenai prediksi kelulusan dengan penerapan algoritma Neural Network berbasis PSO menunjukkan peningkatan akurasi senilai 4,5% dan akurasi sebesar 98,50% dan AUC senilai 0,997%. Selain itu, penerapan metode PSO pada Support Vector Machine juga menunjukkan kenaikan akurasi 0,62% dari 85,81% menjadi 86,43%. Sedangkan pada optimasi menggunakan Particle Swarm Optimization pada Decision Tree mampu memberi kontribusi peningkatan akurasi hingga 87.56% atau dengan kata lain meningkat sebesar 01.01% dibanding tanpa menggunakan optimasi Particle

Swarm Optimization.



## 2.2. Keaslian Penelitian

Tabel 2. 1. Matriks literatur review dan posisi penelitian  
Particle Swarm Optimization Untuk Penentuan Parameter Dan Fitur Seleksi Pada Algoritma K-Nearest Neighbor Pada Prediksi Target Hafalan Siswa SMPIT MIS

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Strategies Memorizing More Rounds May Promote The Emergence Of Cooperation In Stochastic Games	Ting Zou, DKK, Proceedings - IEEE International Symposium on Circuits and Systems, 2021(Scopus)	Untuk menemukan dan mencari metode efisien yang menjaga kerjasama dalam sebuah permainan multi-agen(kelompok).	Ditemukan fakta bahwa melalui pengaturan struktur transisi yang efisien, dapat menghargai kerjasama timbal balik dan menghilangkan pembelotan, beberapa mekanisme yang efisien dapat ditemukan untuk mempromosikan keberlanjutan kerjasama dalam permainan stokastik.	Kedepannya penelitian ini dapat dimanfaatkan untuk merancang metode yang tepat dalam sebuah permainan kelompok	Pada penelitian yang dilakukan oleh Ting Zou DKK. Telah diteliti faktor yang mempengaruhi kerjasama dalam sebuah kelompok, kemudian pada penelitian yang akan dibuat akan dimanfaatkan fakta-fakta yang kemungkinan mempengaruhi faktor kerjasama(kelompok)
2	The Role of Parents in Children Memorizing the Qur'an in Middle School Based on the Amanatul Ummah Islamic Boarding	Mizanul Hasanah, Interdisciplinary Journal of Islamic Education, 2021 (DOAJ)	Bertujuan untuk mengevaluasi hubungan antara orang tua dan anak mengenai kemampuan menghafal Al-Qur-an.	Pada penelitian ini didapatkan fakta bahwa peran orang tua sangat berpengaruh di lingkungan rumah, maka orang tua yang	Penelitian ini kedepannya dapat lebih divarisasikan, seperti mengelompokkan anak	Pada penelitian Mizanul Hasanah telah didapatkan fakta bahwa faktor orang tua sangatlah penting, sehingga

Tabel 2. 1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3	The Effect of Time Management on Academic Achievement among University Students Who Memorizing Qur'an (Hifdz)	Yaumiyyatul Furqoni dan Lilik Mufidah, <i>Advances in Social Science, Education and Humanities Research (ASSEHR)</i> , 2019 (Scopus)	Mencari tahu apakah ketepatan waktu kehadiran mempengaruhi hafalan Al-Qur'an siswa	Faktor pengelolaan waktu, tidak berpengaruh pada objek yang diteliti	Kedepannya dapat dilakukan penelitian dengan objek yang lebih beragam	Pada penelitian yang dilakukan oleh Yaumiyyatul Furqoni dan Lilik Mufidah telah diketahui fakta bahwa faktor ketepatan waktu kehadiran tidak mempengaruhi hafalan siswa, meski begitu pada penelitian ini, akan digunakan untuk mengukur apakah faktor ini berpengaruh pada objek yang diteliti atau tidak
4	Communication Patterns Of Coaches And Students In Cultivating Discipline Of Memorizing The Qur'an	Qudratullah, <i>INJECT (Interdisciplinary Journal of Communication)</i> , 2020 (S2)	Penelitian ini bertujuan untuk mengevaluasi dan mengidentifikasi pola interaksi yang terjadi antara Pembina dan santri dalam menjaga disiplin hafalan Al-Qur'an	Komunikasi antara Guru dan murid terbukti memberi peran penting dalam proses penghafalan Al-Qur'an yang dilakukan oleh siswa	Pada penelitian ini perlu lebih dikelompokkan lagi mengenai kriteria anak yang diteliti, sehingga pengukurannya lebih spesifik	Pada penelitian mengenai pola komunikasi telah didapatkan fakta bahwa faktor komunikasi sangatlah penting, sehingga pada penelitian ini kami akan

Tabel 2. 1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
						memanfaatkan faktor tersebut beserta faktor lainnya yang telah ditemukan.
5	Teori Behavior dalam Meningkatkan Motivasi Hafalan Surat Pendek Al-Quran	Marlina, Jurnal Bimbingan, Penyuluhan, Konseling, dan Psikoterapi Islam, 2018 (S4)	Meneliti apakah ada hubungan antara teori perilaku dan teknik pemberian hadiah dalam meningkatkan semangat belajar hafalan surat pendek Al-Quran	Pemberian reward dan pemberian komunikasi interpersonal yang dilakukan guru sangat efektif dalam meningkatkan motivasi siswa.	Perlu dilakukan pengukuran dengan objek yang lebih beragam	Pada penelitian mengenai teori behavior telah didapatkan fakta bahwa faktor reward sangatlah penting, sehingga pada penelitian ini kami akan memanfaatkan faktor tersebut beserta faktor lainnya yang telah ditemukan.
6	PENGARUH KOMUNIKASI INTERPERSONAL GURU DAN SANTRI TERHADAP MINAT HAFALAN AL-QUR'AN (Studi Pada Santri Putri Takhasus Pondok Pesantren Darul Huffaz	D Oktarina, INTERCODE – Jurnal Ilmu Komunikasi Vol. 1 No. 1 (Maret, 2021), Hal. 49-62 ISSN, 2020 (S2)	Penelitian ini bertujuan untuk mengevaluasi bagaimana interaksi antara pembimbing dan santri dalam meningkatkan motivasi santri putri dalam proses menghafal Al-Qur'an.	Komunikasi interpersonal mampu memberikan dampak signifikan dalam peningkatan minat hafalan Al-Qur'an.	Perlu dibandingkan faktor ini dengan jenis komunikasi lain, sehingga faktor ini tidak mendominasi pengukuran, dan mendapat perbandingan	Pada penelitian komunikasi interpersonal telah didapatkan fakta bahwa faktor komunikasi interpersonal guru dalam memberi motivasi sangatlah penting, sehingga

Tabel 2. 1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	Kecamatan Gedong Tataan Kabupaten Pesawaran Tahun 2020)					pada penelitian ini kami akan memanfaatkan faktor tersebut beserta faktor lainnya yang telah ditemukan.
7	Analisis Prediksi Kelulusan Mahasiswa Menggunakan Decision Tree Berbasis Particle Swarm Optimization	Hendra, Hendra Azis, Mochammad Abdul Suhardjono, Suhardjono, Jurnal Sisfokom (Sistem Informasi dan Komputer), 2020 (S3)	Untuk membantu mahasiswa dalam menentukan kelulusan yang diharapkan dengan meminimalkan jumlah mahasiswa yang tidak lulus tepat waktu melalui cara pemodelan data mining dan machine learning.	Hasil dari penggunaan metode decision tree yang diperbaiki dengan menggunakan algoritma optimasi particle swarm ditunjukkan lebih baik dibandingkan dengan metode decision tree yang digunakan tanpa proses optimasi.	Metode yang digunakan tidak dibandingkan dengan metode lain	Pada penelitian sebelumnya PSO disandingkan dengan DT, namun pada penelitian ini akan diterapkan pada KNN
8	CART and PSO+KNN algorithms to estimate the impact of water level change on water quality in Poyang Lake, China	Li, Yilu, Khan, Mohd Yawar Ali, Jiang, Yunzhong, Tian, Fuqiang, Liao, Weihong, Fu, Shasha, He, Changgao, Arabian Journal of Geosciences, 2019 (Q2)	Untuk mengevaluasi pengaruh ketinggian air di danau terhadap kondisi danau itu sendiri.	Pada periode kalibrasi, rata-rata ketepatan peramalan algoritma CARTs adalah 89,76%, sedangkan hasil algoritma PSO+ KNN memiliki ekspektasi ketepatan rata-rata 95,22%. Konsekuensi prediksi parameter kualitas air CODMn dan	menambahkan variabel yang belum digunakan sebelumnya dan dataset baru. Selanjutnya, perbandingan akan dilakukan dengan algoritma lain untuk mengevaluasi hasilnya.	Terdapat metode yang sama, hanya saja pada penelitian selanjutnya menggunakan semua variabel yang ada, kemudian diseleksi

Tabel 2. 1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				DO berkinerja baik dan tingkat yang benar berada di atas 98,08% (CODMn) dan 95,2% (DO). Untuk parameter kualitas air NH <sub>3</sub> -N dan TP, algoritma PSO+KNN memiliki kemampuan yang lebih dibandingkan algoritma CART.		
9	Model Support Vector Machine Berbasis Particle Swarm	Musyaffa, Nu'man Rifai, Bakhtiar, JITK, 2018 (S3)	Penelitian ini melakukan pengujian untuk memprediksi penyakit liver	Penerapan metode PSO pada Support Vector Machine memiliki hasil yang lebih baik pada akurasi dan nilai AUCnya dibandingkan dengan model Support Vector Machine tanpa penerapan metode tersebut. Perbedaan yang signifikan terlihat pada nilai akurasi sebesar 6% dan nilai AUC sebesar 0.161.	melakukan pengujian menggunakan dataset lain dapat menjadi pertimbangan untuk penelitian kedepannya, hal ini diharapkan dapat menguji model yang telah dilakukan pada penelitian ini	Pada penelitian sebelumnya PSO digunakan untuk meningkatkan performa SVM, namun pada penelitian ini digunakan untuk meningkatkan performa PSO.
10	A novel SVM-kNN- PSO ensemble method	Aburomman, Abdulla Amin	Melakukan deteksi pada intrusi	Penelitian ini menunjukkan bahwa	Penelitian selanjutnya dapat	Terdapat metode yang sama, namun

Tabel 2. 1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	for intrusion detection system	Ibne Reaz, Mamun Bin, Applied Soft Computing Journal, 2016 (Q1)		menggunakan metode PSO dalam ensemble dapat meningkatkan akurasi klasifikasi. Hasil terbaik yang diperoleh dari percobaan menunjukkan peningkatan sebesar 0,756% dibandingkan dengan akurasi dari expert base terbaik. Dengan jumlah data pengujian sebesar 311.029 pengamatan, diharapkan dapat menghasilkan hasil yang lebih baik untuk beberapa pengamatan. Selain itu, metode ini juga diharapkan dapat menyelesaikan tugas dengan waktu yang relatif singkat. Keberhasilan dari ensemble berbasis PSO dapat dikalikan dengan set bobot yang dihasilkan dan	membandingkan berbagai algoritma pengoptimalan untuk menghasilkan bobot.	peneliti akan mencoba mengoptimasinya dengan PSO

Tabel 2. 1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				dioptimalkan untuk mencapai hasil yang lebih baik.		
11	Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation	Meyer, Hanna Reudenbach, Christoph Hengl, Tomislav Katurji, Marwan Nauss, Thomas, Environmental Modelling and Software, 2018 (Q1)	Meningkatkan performa model spatio, bertujuan untuk menunjukkan efek validasi berorientasi target dan menemukan solusi untuk mendeteksi dan mengurangi over-fitting spasial.	Terdapat perbedaan yang signifikan dalam tingkat kesalahan yang diperkirakan antara metode k-fold acak dan metode CV berorientasi target. Dalam kasus studi Tair Antartika, k-fold CV R2 acak yang diperoleh adalah 0,90, sedangkan LLO CV R2 yang diperoleh hanya 0,24. Begitu juga dalam kasus studi VWCookfarm, k-fold CV R2 acak yang diperoleh adalah 0,92, sementara LLO CV R2 yang diperoleh hanya 0,49. Ini menunjukkan bahwa metode k-fold acak dapat memberikan pandangan yang terlalu optimis terhadap hasil,	Melakukan perbandingan lebih banyak mengenai model, untuk menemukan lebih banyak fakta mengenai fungsi dari masing masing algoritma	Terdapat metode yang sama, namun peneliti akan menerapkannya pada algoritma KNN

Tabel 2. 1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				sehingga metode validasi berorientasi target diperlukan untuk menghindari kesalahan perhitungan yang besar.		
12	Comparison of Data Mining Algorithm: PSO-KNN, PSO-RF, and PSO-DT to Measure Attack Detection Accuracy Levels on Intrusion Detection System	Budilaksono, Sularso Riyadi, Andri Agung Azhari, Lukman Saputra, Dedi Dwi Suwanto, M. Anno Agus Suwartane, I. Gede Jupriyanto Ramadhian, Andika Utomo, Agus Prasetyo Fauzi, Achmad, Journal of Physics: Conference Series, 2020 (Q4)	Melakukan perbandingan berbagai algoritma	Pengujian ini menunjukkan bahwa pendekatan yang menggunakan metode PSO-RF mampu menghasilkan akurasi deteksi serangan yang paling tinggi. Hasilnya menunjukkan bahwa algoritma PSO-RF dengan kombinasi jumlah pohon dan kedalaman maksimal - 20 pada dataset CICIDS2017 secara lengkap memiliki akurasi lebih baik dibandingkan dengan algoritma lain yang digunakan. Nilai akurasi tertinggi dicapai pada dataset CICIDS2017	Menggunakan dataset lain dapat menjadi acuan untuk melakukan penelitian selanjutnya	Pada penelitian sebelumnya PSO digunakan untuk berbagai macam algoritma, namun pada penelitian ini PSO hanya akan digunakan untuk KNN

Tabel 2. 1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				secara utuh yaitu 99,76%. Pada dataset CICIDS2017 yang dibagi menjadi 50% Benign dan 50% Attack, algoritma PSO-RF dengan kombinasi jumlah pohon dan maximal depth = 20 juga mendapatkan nilai akurasi tertinggi sebesar 99,67%.		
13.	Optimization of K-nearest neighbor using particle swarm optimization for face recognition	Sasirekha, K., Thangavel, K., Neural Computing and Applications, 2019 (Q1)	Melakukan optimasi pada algoritma untuk kepentingan deteksi wajah	Hasil dari eksperimen menunjukkan bahwa algoritma PSO-KNN mendapatkan hasil lebih baik daripada metode lain yang digunakan untuk pengenalan fitur, seperti tabel keputusan, SVM, MLP, dan KNN. Ini dapat dilihat dari akurasi yang lebih tinggi yang diperoleh dengan jumlah fitur yang lebih sedikit dibandingkan dengan GA dan ACO.	Melakukan pengujian dengan berbagai dataset memungkinkan untuk melihat seberapa efisien penerapan PSO untuk mengoptimasi model yang ada	Terdapat metode yang sama dengan yang peneliti gunakan nantinya, namun Perbedaannya pada jenis kasus yang diteliti

Tabel 2. 1. (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				<p>Oleh karena itu, metode ini dianggap efektif untuk mengenali modalitas biometrik lainnya dan lebih sederhana daripada metode pengenalan wajah CBR, ORB2-IPR, dan UDSPP. Dalam penelitian selanjutnya, varian dari KNN akan dikombinasikan dengan algoritma optimasi lain untuk meningkatkan kinerja dalam pengenalan citra wajah.</p>		

## **2.3. Landasan Teori**

### **2.3.1. Pengumpulan Data**

Subjek atau obyek yang berkualitas serta karakteristik tertentu yang telah ditentukan oleh peneliti digunakan dalam pembentukan wilayah generalisasi yang disebut populasi. Kualitas dan karakteristik ini kemudian diteliti dan ditarik kesimpulannya.

Pemilihan sampel merupakan metode yang digunakan oleh peneliti untuk mengambil sebagian populasi yang nantinya diteliti. Oleh sebab itu, penelitian ini hanya melibatkan sampel yang dianggap representatif dari populasi yang akan diteliti. Hal ini diterapkan sebab terdapat keterbatasan dari sisi tenaga dan waktu serta dana sehingga tidak mungkin untuk meneliti seluruh populasi. Namun, apa yang ditemukan dari sampel tersebut dapat digunakan sebagai representasi dari populasi secara keseluruhan. (Sugiyono, 2016).

Populasi dalam penelitian ini merupakan nilai yang dihasilkan siswa saat belajar yang terdapat pada SMPIT MIS.

### **2.3.2. Data Mining**

Proses data mining merupakan teknik yang dipakai demi mengungkap pola atau informasi penting yang berasal dari data yang besar dan kompleks melalui berbagai tahap, seperti seleksi, pre-processing, transformasi, data mining dan interpretasi evaluasi. Ini menggabungkan teknologi dari empat bidang ilmu komputer, yaitu inteligensi buatan, machine learning, statistik, dan sistem basis data. Hasil dari proses data mining bisa berupa karakteristik data yang deskriptif serta model pengetahuan yang dapat digunakan untuk prediksi. Dari segi

fungsionalitas, data mining telah dibagi 6 menjadi enam kategori, diantaranya klusterisasi, regresi, klasifikasi, deteksi anomali, asosiasi, dan perangkuman (Suyanto, 2017: 1-3).

### 2.3.3. Klasifikasi

Klasifikasi merupakan proses pembagian data ke dalam kelompok yang sudah ditetapkan sebelumnya berdasarkan karakteristik atau fitur tertentu. Melalui metode data mining, sebuah model atau fungsi dapat digunakan untuk memprediksi keanggotaan data tersebut ke dalam suatu grup tertentu dengan mengidentifikasi konsep atau karakteristik dari kelas data tersebut (Muslim et al., 2017: 1). Beberapa metode yang sering dipakai untuk proses klasifikasi meliputi penggunaan algoritma seperti Pohon Keputusan, Mesin Vektor Pendukung, Naive Bayes (NB), dan K-Tetangga Terdekat (Muslim et al., 2018: 141). Klasifikasi data merupakan proses memasukkan data baru ke dalam kelas yang sudah ditentukan sebelumnya. Metode ini dikenal sebagai pembelajaran terawasi karena label/target kelas sudah disediakan sebelumnya. Proses klasifikasi terdiri dari dua tahap utama, yaitu pembelajaran dan klasifikasi. Pembelajaran digunakan untuk membuat pemetaan atau fungsi yang dapat digunakan untuk mengklasifikasikan data baru berdasarkan label kelas yang sudah dikenal. Kemudian, untuk mengukur efektivitas dari classifier, persentase data yang telah diklasifikasikan dengan benar dari semua data yang ada digunakan. Apabila tingkat keakuratan classifier diterima, maka classifier tersebut bisa dipakai mengklasifikasikan data yang label kelasnya belum diketahui. Beberapa algoritma yang sering digunakan dalam klasifikasi di antaranya adalah Decision Tree, SVM, NB, dan KNN. Fungsi target yang diaplikasikan dalam

klasifikasi juga disebut sebagai model klasifikasi. seperti pada Gambar 2.1.



Gambar 2. 1. Blok Diagram Model Klasifikasi

Fungsi target diketahui sebagai model klasifikasi. Menurut Hermawati (2013: 56) terdapat dua jenis model klasifikasi diantaranya:

- a. Pemodelan Deskriptif, merupakan proses untuk membuat suatu model yang dapat digunakan sebagai alat untuk menjelaskan perbedaan objek yang ada pada kelas yang berbeda.
- b. Pemodelan prediktif, merupakan proses untuk membuat suatu model dan digunakan untuk mengklasifikasikan data yang label kelasnya belum diketahui. Algoritma pembelajaran digunakan untuk membuat model yang mencerminkan hubungan antara label kelas dan atribut dalam data masukan. Tujuan proses ini membuat suatu model yang mampu memprediksi label kelas dari data yang belum diketahui dengan tingkat akurasi yang tinggi (Hermawati, 2013: 56).

#### 2.3.4. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) merupakan algoritma untuk proses klasifikasi data. Algoritma ini termasuk dalam kategori pembelajaran berbasis instansi dan pembelajaran malas. KNN bekerja mencari kelompok pada  $n$  objek dari data pelatihan yang sangat serupa dengan objek dalam data uji atau data baru. Klasifikasi objek dilakukan berdasarkan jarak terdekat dari data pelatihan yang ada. Dalam KNN, objek diklasifikasikan berdasarkan jarak dengan tetangga

terdekat, dengan objek yang ditugaskan ke kelas yang paling banyak di antara tetangga terdekatnya. Biasanya, jarak yang digunakan adalah jarak Euclidean (Karegowda et al., 2012: 148):

1. Menentukan nilai  $K$ .
2. Data testing dan data training dihitung jarak Euclidean.
3. Jarak Euclidean menjadi landasan dalam mengelompokkan data.
4. Nilai terkecil/tetangga terdekat menjadi patokan dalam mengelompokkan data.
5. Pemilihan hasil prediksi dengan kriteria kelas yang paling sering muncul dari sejumlah  $K$ .

Data latih dalam atribut ke-1 diperoleh pada Persamaan 1.

$$X1 = (X11, X12, \dots, X1n) \quad (1)$$

Data latih dalam atribut ke-2 diperoleh pada Persamaan 2.

$$X2 = (X21, X22, \dots, X2n) \quad (2)$$

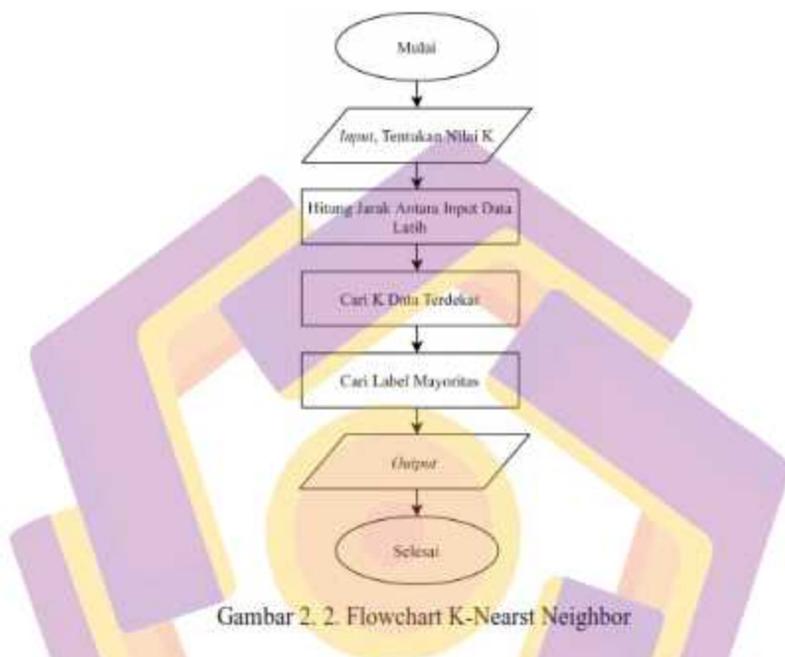
jarak Euclidean diperoleh pada Persamaan 3.

$$d(X_1, X_2) = \sqrt{\sum_r^n (a_r(X_1) - a_r(X_2))^2} \quad (3)$$

$X1$  dan  $X2$  merupakan dua record dengan jumlah  $n$  atribut, Perhitungan sela antara dua record  $X1$  dan  $X2$  dilakukan demi mengetahui seberapa jauh perbedaan nilai antar atribut pada record tersebut. Data baru diklasifikasikan dengan menghitung sela antara data baru dengan record yang ada pada dataset menggunakan metode Euclidean distance. Dari hasil perhitungan jarak tersebut, akan dipilih  $k$  record terdekat untuk digunakan dalam proses klasifikasi. Kemudian, dari  $k$  record tersebut, kelas yang paling banyak muncul akan

digunakan sebagai keputusan akhir dalam pengklasifikasian data baru.

Proses kerja KNN dapat dilihat pada Gambar 2.2.



### 2.3.5. Forward Selection

Forward Selection (FS) merupakan metode yang digunakan untuk mengevaluasi keefektifan setiap atribut dalam data dan menghilangkan atribut tidak relevan (J. Han dan M. Kamber, 2006). Algoritma FS berpatokan dengan model regresi linear (R. Noori, dkk., 2011).

Forward selection adalah salah satu metode pemilihan fitur dalam machine learning, dimana fitur-fitur yang akan digunakan dalam model dipilih satu per satu, Forward selection juga merupakan salah satu metode untuk membangun model K-Nearest Neighbor (KNN) secara bertahap dengan menambahkan fitur

satu per satu. Berikut ini adalah contoh perhitungan manual forward selection pada KNN:

1. Langkah pertama adalah menentukan nilai  $K$  pada KNN. Misalkan  $K = 3$ .
2. Selanjutnya, kita harus memilih fitur yang paling penting dalam memprediksi target variable. Misalkan kita memiliki dataset dengan 3 fitur, yaitu fitur  $A$ ,  $B$ , dan  $C$ . Untuk memilih fitur yang paling penting, kita dapat menggunakan metode pengujian satu per satu (one-at-a-time testing). Dalam metode ini, kita akan membuat model KNN dengan hanya satu fitur, dan memilih fitur yang memberikan akurasi prediksi tertinggi.
  - a. Kita memulai dengan membangun model KNN dengan hanya menggunakan fitur  $A$ . Untuk memilih nilai  $K$ , kita dapat menggunakan metode pengujian silang (cross-validation) dengan membagi dataset menjadi  $k$ -fold dan menghitung akurasi model pada setiap fold. Misalkan hasil pengujian silang menunjukkan bahwa model KNN dengan fitur  $A$  dan  $K = 3$  memiliki akurasi 80%.
  - b. Selanjutnya, kita membangun model KNN dengan menggunakan fitur  $A$  dan  $B$ . Dalam metode forward selection, kita harus mempertimbangkan fitur tambahan secara berurutan. Oleh karena itu, kita harus membangun model dengan fitur  $A$  terlebih dahulu, baru kemudian menambahkan fitur  $B$ . Misalkan hasil pengujian silang menunjukkan bahwa model KNN dengan fitur  $A$  dan  $B$ , serta  $K = 3$  memiliki akurasi 85%.

c. Berikutnya, kita membangun model KNN dengan menggunakan fitur A, B, dan C. Misalkan hasil pengujian silang menunjukkan bahwa model KNN dengan fitur A, B, dan C, serta  $K = 3$  memiliki akurasi 87%.

3. Setelah melakukan pengujian pada setiap kombinasi fitur, kita memilih model KNN dengan akurasi tertinggi sebagai model terbaik. Dalam contoh ini, model KNN dengan fitur A, B, dan C serta  $K = 3$  memiliki akurasi tertinggi, sehingga dipilih sebagai model terbaik untuk memprediksi target variable.

### 2.3.6. Particle Swarm Optimization

Menurut Chen (2013: 230), tahapan PSO diawal ditentukan kecepatan awal dan posisi awal secara acak.

Selanjutnya, proses pengembangan algoritma ini dilakukan dengan cara sebagai berikut:

1. Dalam proses optimisasi menggunakan PSO, diasumsikan bahwa jumlah partikel dalam kawanan (atau ukuran kelompok) adalah  $N$ . Setiap partikel dalam  $N$  dimensi diinisialisasi dengan posisi dan kecepatan yang acak. Kecepatan awal dari semua partikel dianggap nol dan proses iterasi dimulai dari 1. Setiap partikel akan bergerak menuju titik optimal dengan kecepatan yang ditentukan.
2. Untuk menghitung fitness dari partikel, dapat dilakukan dengan menggunakan rumus fungsi fitness berikut:

$$Fitness = Error\ rate = \frac{\text{jumlah instance yang diklasifikasikan salah}}{\text{jumlah instance}} \quad (4)$$

3. Pbest dan gbest diperbaharui berdasarkan fungsi fitness:

$$\text{If } (pos > pbest): pbest = pos \quad (5)$$

$$\text{If } (pos > gbest): gbest = pos$$

4. Perbarui kecepatan dengan Persamaan berikut:

$$v_{id}^{new} = w \cdot v_{id}^{old} + c_1 \cdot r_1 (pb_{id}^{old} - x_{id}^{old}) + c_2 + r_2 (gb_{id}^{old} - x_{id}^{old}) \quad (6)$$

Keterangan:

$w$  = Parameter inertia weight (kontrol perilaku konvergensi PSO)

$r_1$  dan  $r_2$  = Parameter acak (random) 0-1

$c_1$  dan  $c_2$  = Konstanta akselerasi (learning rate), menandakan gerakan iterasi partikel kontrol

$x_{id}^{old}$  = Letak awal partikel  $i$  di  $d$  dimensi

$gb_{id}^{old}$  = gbest (global best) di  $d$  dimensi

$pb_{id}^{old}$  = pbest (personal best) partikel  $i$  di  $d$  dimensi

5. Perbarui posisi dengan Persamaan

$$x_{id}^{new} = x_{id}^{old} + v_{id}^{new} \quad (7)$$

Dimana:

$x_{id}^{old}$  = Posisi awal partikel  $i$  pada  $d$  dimensi

$v_{id}^{new}$  = Kecepatan baru partikel  $i$  pada  $d$  dimensi

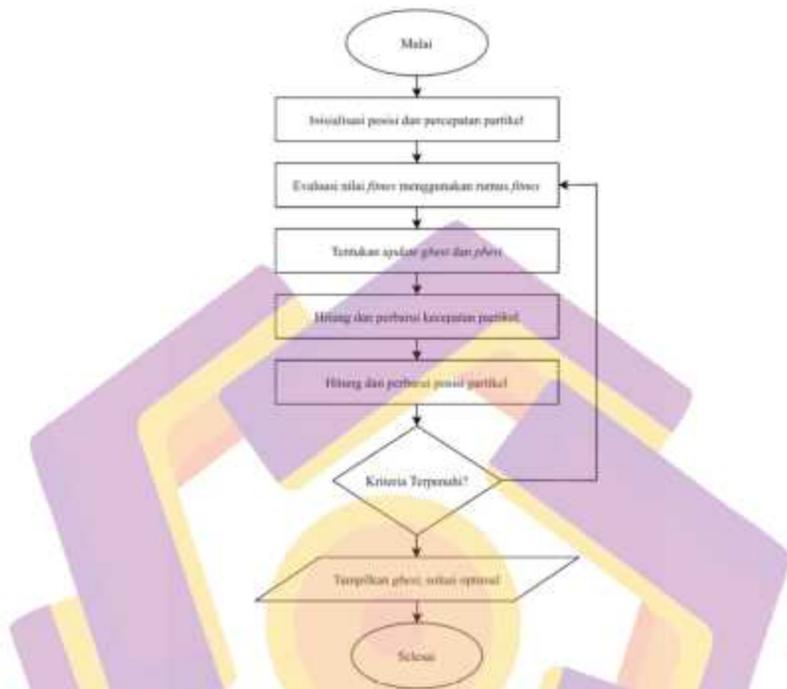
6. Untuk menentukan apakah proses telah mencapai konvergensi, dapat dilakukan dengan mengecek posisi partikel apakah telah berada pada nilai yang sama. Jika ini terjadi, proses dianggap telah konvergen. Namun, jika belum, iterasi diperbarui dengan mengubah  $i$  menjadi  $i+1$  dan memperbarui nilai pbest dan gbest. Proses ini diulang sampai semua

partikel berada pada solusi yang sama. Biasanya, kriteria penghentian ditentukan, seperti ketika perbedaan solusi saat ini dibandingkan solusi sebelumnya sangat rendah.

7. Menurut vandenBergh & Engelbrecht (2004: 232), terdapat dua hal penting yang harus dipertimbangkan ketika kondisi penghentian (stopping criteria) dipilih dalam proses PSO, yaitu:

- a. Kondisi penghentian tidak menyebabkan proses konvergensi yang terlalu cepat (premature convergence), di mana solusi yang diperoleh tidak optimal.
- b. Kondisi penghentian harus diterapkan dengan hati-hati agar terhindar dari masalah oversampling, yaitu jika proses penghentian mengharuskan perhitungan yang berkelanjutan, maka akan menambah kompleksitas dari proses optimisasi.

Tahapan-tahapan dari algoritma PSO terlihat pada Gambar 2.3.



Gambar 2. 3. Tahapan algoritma Particle Swarm Optimization

### 2.3.7. Confusion Matrix

Confusion matrix (Kohavi & Provost, 1998) digunakan untuk menyajikan informasi mengenai klasifikasi aktual dan prediksi yang dilaksanakan oleh sistem klasifikasi, khususnya untuk dua kelas klasifikasi. Elemen-elemen dari confusion matrix digunakan menghitung sensitivitas, spesifisitas, akurasi klasifikasi, nilai prediktif negatif, dan nilai prediktif positif (Akay, 2008: 3243-3244).

Tabel 2. 2. Representasi Confusion Matrix

Actual	Predicted	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Akurasi merupakan persentase dari total data yang dinilai dan diidentifikasi.

Rumus perhitungan akurasi ditunjukkan pada Persamaan 8.

$$(\%) = \frac{TP + TN}{TP + FP + FN + TN} \quad (8)$$

Rumus Precision ditunjukkan pada persamaan 9.

$$precision = \frac{TP}{TP + FP} \quad (9)$$

Rumus Recall ditunjukkan pada persamaan 10.

$$recall = \frac{TP}{TP + FN} \quad (10)$$

## **BAB III**

### **METODE PENELITIAN**

#### **3.1. Jenis, Sifat dan Pendekatan Penelitian**

Jenis penelitian yang diterapkan oleh peneliti merupakan penelitian Eksperimen dimana peneliti melakukan perhitungan matematis untuk menemukan hasil yang diinginkan, metode ini diterapkan oleh peneliti untuk mengidentifikasi pengaruh variabel bebas yang dipakai sebagai perlakuan terhadap variabel terikat disertai kondisi yang terkontrol.

Sifat dari penelitian yang akan dilakukan adalah Kausal, dimana peneliti melakukan pengujian untuk mencari tahu karakteristik variabel dalam mempengaruhi akurasi terhadap suatu model agar dapat lebih maksimal dalam mengukur tingkat prediksi target hafalan siswa, penelitian ini mendeteksi performa penerapan algoritma KNN yang ditingkatkan menggunakan PSO dan fitur seleksi data.

Pendekatan kuantitatif diaplikasikan, maka penelitian akan diproses sesuai alur yang telah ditentukan.

#### **3.2. Metode Pengumpulan Data**

Data yang dipakai pada analisis ini berasal dari data siswa SMPIT MIS berupa nilai-nilai pelajaran dalam satu semester dan merupakan data primer yang dimana data ditarik dan diolah dari SMPIT MIS kemudian disatukan dan dikonversi kedalam angka, lalu variabel lainnya ialah Faktor Kerjasama kelompok, peran orang tua, disiplin waktu, pelaksanaan dan evaluasi, pola komunikasi guru dan

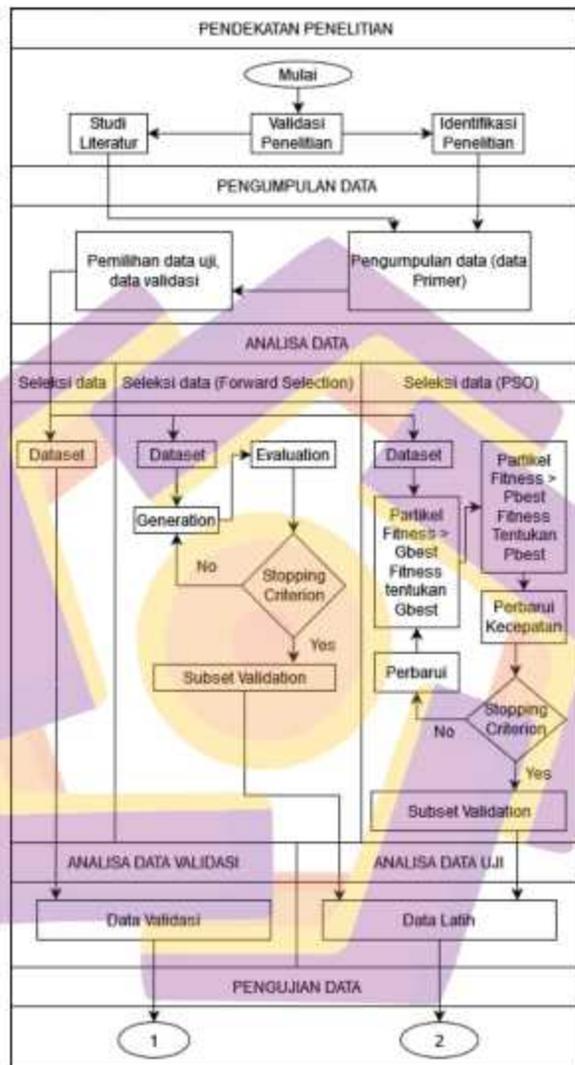
siswa, pemberian motivasi, reward, minat siswa, perubahan lingkungan internal dan eksternal yang dapat diukur, usia, lokasi, dan manajemen waktu, serta faktor sikap dan kepribadian, selain itu nilai baca dan tulis Al-Qur-an, kemampuan beradaptasi dengan metode menghafal, serta motivasi atau tingkat kerajinan siswa dalam menyeter hafalan mereka, kemudian nilai-nilai lainnya seperti keaktifan siswa dalam mengerjakan amalan-amalan seperti sholat, mengaji, dan puasa. Jumlah data yang digunakan adalah 63 data siswa, karakteristik data berupa angka dengan nilai terendah adalah 0 sedangkan tertinggi adalah 100. Data ini merupakan data populasi siswa SMPIT MIS mulai kelas 7 hingga kelas 9, nilai pada siswa didapatkan dari hasil belajar selama program karantina Al-Qur'an yang kemudian data ini yang akan diolah nantinya.

### **3.3. Metode Analisis Data**

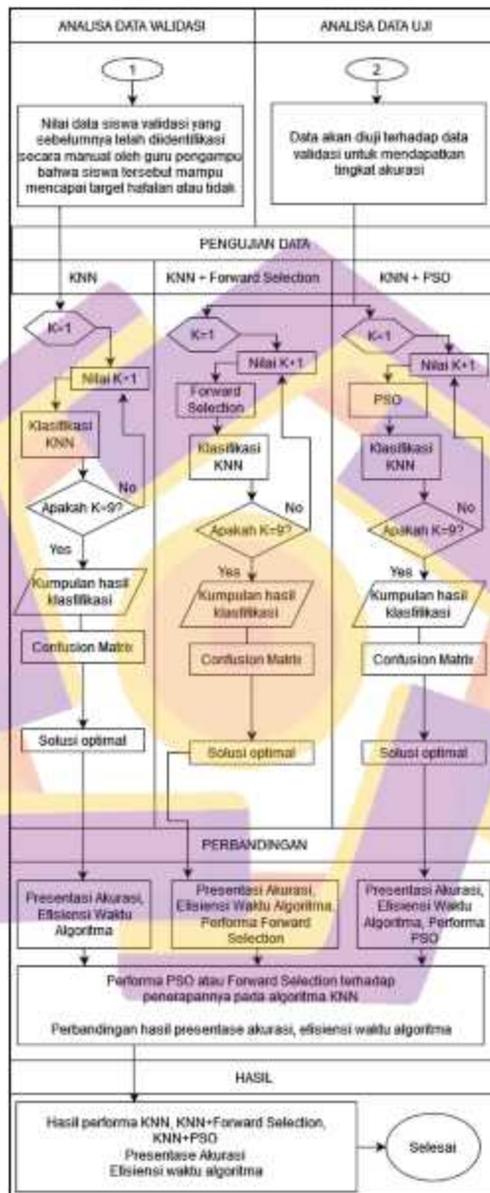
Analisis data dilakukan menggunakan teknik seleksi data yang dimana data diseleksi sebelum dilakukan pemrosesan ke tahap selanjutnya, menggunakan algoritma FS dan PSO, hal ini untuk menyeleksi atribut yang tidak memiliki pengaruh besar dalam akurasi algoritma KNN, kemudian untuk mengukur percobaan yang dilakukan, maka digunakan confusion matrix.

### **3.4. Alur Penelitian**

Alur penelitian yang diaplikasikan pada penelitian ini ditunjukkan pada Gambar 3.1. dan Gambar 3.2. sebagai berikut :



Gambar 3. 1. Diagram alur penelitian bagian 1



Gambar 3. 2. Diagram alur penelitian bagian 2

Berikut penjelasan pada alur penelitian pada Gambar 3.1. sebagai berikut:

Pendekatan Penelitian (Validasi Penelitian, Studi Literatur, Identifikasi Masalah). Tahap awal penelitian yaitu melakukan validasi penelitian tentang permasalahan serta studi literatur untuk mengetahui masalah yang akan di analisa serta penggunaan algoritma yang tepat berdasarkan studi literatur sehingga dalam pemilihan algoritma tersebut didasarkan pada hasil ilmu yang dianalisa oleh penelitian sebelumnya.

1. Pendekatan Penelitian (Validasi Penelitian, Studi Literatur, Identifikasi Masalah) Tahap awal penelitian yaitu melakukan validasi penelitian tentang permasalahan serta studi literatur untuk mengetahui masalah yang akan di analisa serta penggunaan algoritma yang tepat berdasarkan studi literatur sehingga dalam pemilihan algoritma tersebut didasarkan pada hasil ilmu yang dianalisa oleh penelitian sebelumnya.

## 2. Metode Pengumpulan Data (Pemilihan Data Uji, Data Validasi)

Dalam tahap pengambilan data, peneliti melakukan upaya-upaya yang diperlukan untuk memperoleh data yang bisa diandalkan keakuratannya, agar informasi yang didapatkan bisa dipertanggungjawabkan kebenarannya. Jenis data yang diaplikasikan sumbernya berasal dari data primer yang dimana peneliti mendapatkan data dan diolah terlebih dahulu sebelum digunakan. Data yang dipakai pada analisis ini berasal dari data siswa SMPIT MIS berupa nilai-nilai pelajaran dalam satu semester, kemudian disatukan dan dikonversi kedalam angka, lalu variabel lainnya yang berjumlah total keseluruhan ada 18 variabel diolah hingga membentuk dataset. Jumlah data yang digunakan adalah 63 baris

data, karakteristik data berupa angka dengan nilai terendah adalah 1 sedangkan tertinggi adalah 100. Data ini merupakan data populasi sisa SMPIT MIS mulai kelas 7 hingga kelas 9.

### 3. Analisa Data menggunakan FS, PSO, dan tanpa melalui preprocessing data

Tahap ini dilakukan untuk memperkecil jumlah data yang digunakan, salah satunya dengan menggunakan teknik attribute subset selection. Teknik ini digunakan untuk mengurangi jumlah atribut yang tidak relevan atau yang duplikat. Pada penelitian ini, teknik seleksi atribut yang digunakan meliputi Forward Selection, Particle Swarm Optimization (PSO), dan tanpa melalui preprocessing data atau dengan kata lain data langsung dimasukkan ke KNN untuk dihitung akurasi.

### 4. Tahapan Pembagian Data Training dan Data Testing

Pembagian data diterapkan metode splitter yang secara otomatis memisahkan data dengan perbandingan sebesar 80:20 pada data training dan data testing. Pembagian secara acak dengan mengaplikasikan konsistensi pengacakan (random state) tertentu, sehingga setiap eksekusi hasilnya akan tetap sama.

### 5. Pengujian data

Pada tahap ini dataset yang telah siap diolah dilakukan pengujian menggunakan tiga metode yaitu menggunakan KNN, menggunakan KNN + FS, dan menggunakan KNN + PSO

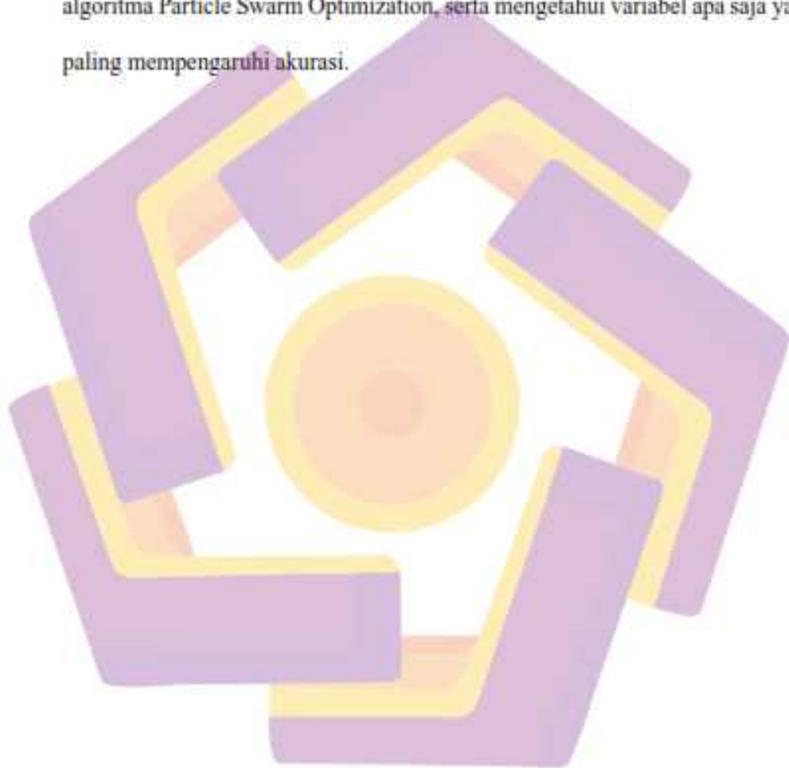
### 6. Tahap Evaluasi menggunakan Confusion Matrix

Tahap evaluasi dilakukan setelah proses penelitian selesai untuk menguji keakuratan model yang telah diterapkan. Teknik yang aplikasikan untuk

mengevaluasi model adalah dengan menggunakan confusion matrix.

#### 7. Dokumentasi Hasil Penelitian

Hasil pada penelitian ini untuk mengetahui hasil performa KNN setelah dilakukan peningkatan dengan menggunakan algoritma Forward Selection atau algoritma Particle Swarm Optimization, serta mengetahui variabel apa saja yang paling mempengaruhi akurasi.



## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

#### 4.1. Pengolahan Data

Data hasil belajar karantina Al-Qur'an siswa SMPIT MIS yang memiliki 18 variabel serta 63 baris data diolah hingga siap untuk diproses, data dengan 18 variabel ini merupakan faktor yang dapat mempengaruhi target hafalan siswa yang telah dikumpulkan berdasarkan penelitian terdahulu yang kemudian nantinya akan dilakukan penyeleksian sehingga atributnya berkurang, hal ini dilakukan agar pada proses pengelompokan siswa pada program karantina Al-Qur'an bisa lebih mudah pembagiannya karena atribut penilaiannya semakin sedikit. adapun contoh datanya adalah

Tabel 4. 1. Dataset siswa

No.	NAMA	NILAI SEMESTER	KERJASAMA	...	STATUS KELULUSAN
1	ALYSHA	90	90		LULUS
2	NISRINA	90	90		LULUS
3	NURUL	70	90		LULUS
...	...	...	...	...	...
63	ARINI	93	30		LULUS

Data ini terdiri dari 37 data yang status kelulusannya adalah lulus dan 26 data yang status kelulusannya adalah tidak lulus, kemudian pada tahap berikutnya data nama dan status kelulusan pada dataset dihapuskan,

Kemudian data yang ada dibagi menjadi dua bagian yaitu data latih dan data uji, data ini dibagi dengan perbandingan 80:20, pada tahapan ini digunakan random state dengan nilai 99 (hal ini agar hasil akurasi yang didapatkan tidak berubah-ubah

setiap kali proses dijalankan dan merupakan angka random). Adapun contoh data untuk memudahkan maka tampilan variabel diubah menjadi angka adapun tampilannya sebagai berikut.

Tabel 4. 2. Data latih

0	1	2	3	4	5	6	7	8	9	10	11	...	17
77	90	35	95	85	70	75	90	25	75	75	30	...	1
90	90	90	89	85	85	85	90	90	85	95	93	...	1
92	90	45	90	85	70	75	90	90	30	75	30	...	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...
88	30	20	80	85	85	75	90	89	30	60	20	...	1

Pada data latih, terdapat total 50 baris data

Tabel 4. 3. Data Uji

0	1	2	3	4	5	6	7	8	9	10	11	...	17
93	30	45	88	85	70	75	90	90	30	75	20	...	1
90	90	90	85	85	85	85	90	85	85	95	93	...	2
89	90	40	89	85	85	85	90	85	30	90	30	...	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
83	90	89	95	85	85	85	90	88	85	30	86	...	1

Pada data uji terdapat total 13 baris data

#### 4.2. Penerapan Algoritma K-Nearest Neighbor, Forward Selection, dan Particle Swarm Optimization (PSO)

Penerapan algoritma KNN, FS, dan PSO dilakukan dengan memakai bahasa pemrograman Python dengan memanfaatkan Notebook Google Colab dengan bantuan library scikit-learn untuk mengimplementasikan algoritma KNN, dan FS, sedangkan untuk algoritma PSO menggunakan berbagai library seperti numpy, pandas, matplotlib, dan seaborn.

#### 4.2.1. Penerapan Algoritma KNN

Pada bagian ini akan dibahas mengenai langkah proses dari analisa data menggunakan FS dan PSO, yang nantinya dievaluasi menggunakan KNN, dimana pada penelitian ini diterapkan nilai  $K=2,3,4,5,6,7,8$ . Perhitungan ini menggunakan library sklearn. Selanjutnya untuk metode KNN+FS dan KNN+PSO akan mencari K optimal yang digunakan untuk memperbaiki kelemahan pada metode KNN dengan menyeleksi variabel yang ada.

Dalam contoh penerapan metode KNN, K yang digunakan adalah  $K=2,3,4,5,6,7,8$  tetapi dalam contoh penerapan ini peneliti hanya memaparkan langkah kerja pada saat jumlah  $K=3$  menggunakan library sklearn.

1. Menentukan nilai K,  $K=3$ .
2. Kemudian menghitung euclidean distance, pada dataset yang telah dibagi menjadi data uji dan data latih telah menghasilkan data latih sebanyak 50 baris dan data uji sebanyak 13 baris.

Contoh perhitungan data uji yang pertama adalah, NILAI SEMESTER=77, KERJASAMA=90, PERAN ORANGTUA=35, DISIPLIN=95, PELAKSANAAN EVALUASI=85, POLA KOMUNIKASI=70, PEMBERIAN MOTIVASI=70, TINGKAT=75, BTQ=90, KEAKTIFAN=25, KEPRIBADIAN=75, PERUBAHAN LINGKUNGAN=75, MINAT=30, REWARD=75, ADAPTASI=30, MOTIVASI=95, UMUR=12, JENIS KELAMIN=1.

Dan data pelatihan yang pertama NILAI SEMESTER=93, KERJASAMA=30, PERAN ORANGTUA=45, DISIPLIN=88, PELAKSANAAN EVALUASI=85, POLA KOMUNIKASI=70, PEMBERIAN MOTIVASI=75, TINGKAT=90,

BTQ=90, KEAKTIFAN=30, KEPRIBADIAN=75, PERUBAHAN LINGKUNGAN=20, MINAT=30, REWARD=87, ADAPTASI=30, MOTIVASI=85, UMUR=14, JENIS KELAMIN=1.

Dengan contoh perhitungan menggunakan rumus adalah:

$$d1 = \sqrt{(77-93)^2 + (90-30)^2 + \dots + (12-14)^2 + (1-1)^2} = 111,7318218$$

Selanjutnya dilakukan perhitungan euclidean distance dengan semua data latih.

Tabel 4. 4. Hasil perhitungan data uji pertama dengan semua data pelatihan

No.	Jarak
1	111,7318218
2	6,480740698
3	24,73863375
4	83
5	83,91662529
...	...
50	28,46049894

3. Data uji pertama dihitung euclidean distancenya dengan semua data latih. Berikut hasil metode KNN dari jarak terkecil ke besar.

Hasil pengurutan perhitungan data uji pertama dengan semua data pelatihan metode KNN

Tabel 4. 5. Hasil pengurutan perhitungan data uji pertama dengan semua data pelatihan metode KNN

No.	Jarak
2	6
30	17

Tabel 4. 5. (Lanjutan)

No.	Jarak
3	24
50	28
37	52
...	...
22	133

4. Mengambil nilai terkecil sesuai nilai K yaitu 3. Berikut diperlihatkan 3 jarak terdekat data uji terhadap data pelatihan.

Tabel 4. 6. Jarak terdekat sesuai nilai K Metode K-NN

No.	Jarak	Label
2	6	Lulus
30	17	Lulus
3	24	Lulus

5. Label mayoritas menjadi klasifikasi

Dari Tabel 4.6 diperoleh data klasifikasi kategori Lulus = 3, dan klasifikasi kategori Tidak Lulus = 0, sehingga hasil klasifikasi dengan  $K = 3$  adalah Lulus. Berikut label aktual dan label prediksi hasil klasifikasi data uji metode KNN

Tabel 4. 7. Label aktual dan label prediksi hasil klasifikasi semua data uji metode K-NN

No.	Label Aktual	Label Prediksi
1	Lulus	Lulus

Tabel 4. 7. Lanjutan)

No.	Label Aktual	Label Prediksi
2	Lulus	Lulus
3	Tidak Lulus	Tidak Lulus
4	Lulus	Tidak Lulus
5	Lulus	Tidak Lulus
6	Tidak Lulus	Tidak Lulus
7	Tidak Lulus	Lulus
8	Tidak Lulus	Lulus
9	Tidak Lulus	Lulus
10	Tidak Lulus	Lulus
11	Tidak Lulus	Lulus
12	Lulus	Lulus
13	Lulus	Lulus

#### 6. Hitung akurasi metode KNN

Dari Tabel 3.4 TP = 4, TN = 2, FP= 2, FN= 5. Terdapat 13 data uji, dengan 2 label yaitu Lulus dan Tidak Lulus. Berikut hasil klasifikasi Metode KNN menggunakan confusion matrix

Tabel 4. 8. Confussion Matrix KNN

	True LULUS	True TIDAK LULUS	Class precision
pred LULUS	4 (TP)	2 (FP)	71%
Pred. TIDAK LULUS	2 (FN)	5 (TN)	67%
Class recall	71%	67%	

Keterangan tabel 4.8. adalah:

- a. Banyaknya data yang seharusnya LULUS kemudian diprediksi LULUS sebanyak 4.
- b. Banyaknya data yang seharusnya TIDAK LULUS kemudian diprediksi TIDAK LULUS sebanyak 5.
- c. Banyaknya data yang seharusnya TIDAK LULUS kemudian diprediksi LULUS sebanyak 2.
- d. Banyaknya data yang seharusnya LULUS kemudian diprediksi TIDAK LULUS sebanyak 2.

Evaluasi pada model mengaplikasikan nilai akurasi dan classification error.

Akurasi dan classification error dari model yaitu :

$$Akurasi = \frac{4+6}{4+2+2+5} = 69.00\%$$

$$Error = \frac{2+2}{4+2+2+5} = 21.00\%$$

7. Pada penelitian ini menggunakan dataset karantina hasil belajar SMPIT MIS, dengan dataset sebanyak 63 baris data, proses perhitungan menggunakan KNN dengan jumlah nilai  $K=3$ , menghasilkan akurasi sebesar 69.00%.

8. Setelah itu, dilakukan perulangan dari range 2 sampai 8 untuk mengetahui nilai  $K$  terbaik sekaligus menghitung akurasi dari setiap nilai  $K$ . hasil perhitungan akurasi dari setiap nilai  $K$  seperti pada Gambar 4.1.

```

for i in range(2, 9):
    knn = KNeighborsClassifier(n_neighbors=i)
    knn.fit(X_train, y_train)
    print(f"k-value = , {i}")
    print(knn.score(X_test, y_test))
    print(f"-----")

```

```

k-value = , 2
0.6153846153846154
-----

```

```

k-value = , 3
0.6923076923076923
-----

```

```

k-value = , 4
0.5384615384615384
-----

```

```

k-value = , 5
0.6923076923076923
-----

```

```

k-value = , 6
0.6923076923076923
-----

```

```

k-value = , 7
0.6923076923076923
-----

```

```

k-value = , 8
0.6923076923076923
-----

```

Gambar 4. 1. Akurasi K value KNN

berikut hasil perhitungan target hafalan siswa menggunakan KNN, pada proses ini nilai K mengalami perubahan dari 2 hingga 8 seperti pada tabel 4.9.

Tabel 4. 9. Hasil Perhitungan KNN

Nilai K	Akurasi
2	61.53%
3	69.23%
4	53.84%
5	69.23%

Tabel 4. 9. (Lanjutan)

Nilai K	Akurasi
6	69.23%
7	69.23%
8	69.23%

Pada tabel di atas, diketahui bahwa nilai akurasi terbaik adalah ketika K bernilai 3 yaitu mencapai 69.23%, sedangkan setelah nilai K ditingkatkan akurasi tidak mengalami peningkatan sehingga didapatkan nilai terbaik adalah ketika K bernilai 3. Seperti pada Gambar 4.12. Precision, recall, f1-score, dan lainnya disebut dengan metrik evaluasi model machine learning atau metrik akurasi. Metrik akurasi digunakan untuk menilai performa model machine learning dalam mengelompokkan sesuatu ke dalam kelas yang benar.

```
print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
0	0.67	0.67	0.67	6
1	0.71	0.71	0.71	7
accuracy			0.69	13
macro avg	0.69	0.69	0.69	13
weighted avg	0.69	0.69	0.69	13

Gambar 4. 2. Metrik evaluasi KNN

Metrik akurasi digunakan untuk mengevaluasi kinerja model pembelajaran mesin dengan membandingkan hasil prediksi model dengan data aktual. Metrik akurasi yang paling umum digunakan adalah precision, recall, dan f1-score.

Berdasarkan hasil metrik evaluasi KNN pada Gambar 4.2. dapat disimpulkan beberapa hal berikut:

1. Model memiliki nilai precision sebesar 0.67 untuk kelas 0 dan 0.71 untuk kelas 1. Ini berarti model kurang baik dalam mengelompokkan sesuatu ke dalam kelas yang benar, namun masih ada kemungkinan model terkadang benar mengelompokkan sesuatu.
2. Model memiliki nilai recall sebesar 0.67 untuk kelas 0 dan 0.71 untuk kelas 1. Ini berarti model kurang baik dalam menemukan semua contoh yang benar dari kelas 0 dan kelas 1, namun masih ada beberapa contoh yang benar yang ditemukan oleh model.
3. Model memiliki nilai f1-score sebesar 0.67 untuk kelas 0 dan 0.71 untuk kelas 1. Ini berarti model kurang baik dalam menemukan contoh yang benar dari kelas 0 dan kelas 1, namun masih ada beberapa contoh yang benar yang ditemukan oleh model.
4. Model memiliki nilai accuracy sebesar 0.69. Ini berarti model kurang baik dalam mengelompokkan semua contoh ke dalam kelas yang benar, namun masih ada kemungkinan model terkadang benar mengelompokkan sesuatu.
5. Model memiliki nilai macro avg sebesar 0.69. Ini berarti model memiliki nilai precision, recall, dan f1-score yang kurang baik untuk semua kelas.

6. Model memiliki nilai weighted avg sebesar 0.69. Ini berarti model memiliki nilai precision, recall, dan f1-score yang kurang baik untuk semua kelas, dihitung dengan memperhitungkan jumlah contoh dari masing-masing kelas (support).

Secara keseluruhan, model memiliki performa yang kurang baik dalam mengelompokkan sesuatu ke dalam kelas yang benar. Ada beberapa kekurangan yang perlu diperbaiki agar model dapat lebih baik dalam mengelompokkan sesuatu ke dalam kelas yang benar.

#### 4.2.2. Penerapan Algoritma FS pada KNN

Forward selection adalah salah satu metode untuk membangun model K-Nearest Neighbor (KNN) secara bertahap dengan menambahkan fitur satu per satu. Berikut ini adalah contoh perhitungan manual forward selection pada KNN:

1. Langkah pertama adalah menentukan nilai K pada KNN. Misalkan  $K = 3$ .

Selanjutnya, kita harus memilih fitur yang paling penting dalam memprediksi target variable. Misalkan kita memiliki dataset dengan 3 fitur, yaitu fitur A, B, dan C. Untuk memilih fitur yang paling penting, kita dapat menggunakan metode pengujian satu per satu (one-at-a-time testing).

Pada penelitian kali ini digunakan 18 variabel, yang kemudian nantinya masing-masing variabel dimasukkan satu persatu lalu dihitung akurasi, berikut 18 variabel yang digunakan

```
# iterating the columns
for col in data.columns:
    print(col)
```

```
NAMA
NILAI SEMESTER
KERJASAMA
PERAN ORANGTUA
DISIPLIN
PELAKSANAAN EVALUASI
POLA KOMUNIKASI
PEMBERIAN MOTIVASI
TINGKAT
BTQ
KEFAKTIFAN
KEPRIBADIAN
PERUBAHAN LINGKUNGAN
HINAT
REWARD
ADAPTASI
MOTIVASI
UMUR
JENIS KELAMIN
STATUS KELULUSAN
```

Gambar 4. 3. Menelusuri nama kolom dari dataset

Yang kemudian dataset diperiksa apakah ada missing valuenya

```
# checking missing values in the data
data.isnull().sum()
```

```
NAMA          0
NILAI SEMESTER  0
KERJASAMA     0
PERAN ORANGTUA  0
DISIPLIN     0
PELAKSANAAN EVALUASI  0
POLA KOMUNIKASI  0
PEMBERIAN MOTIVASI  0
TINGKAT      0
BTQ          0
KEFAKTIFAN   0
KEPRIBADIAN  0
PERUBAHAN LINGKUNGAN  0
HINAT        0
REWARD       0
ADAPTASI     0
MOTIVASI     0
UMUR         0
JENIS KELAMIN  0
STATUS KELULUSAN  0
dtype: int64
```

Gambar 4. 4. Menampilkan semua fitur dan mengecek missing value

Pada penelitian ini variabel yang ada telah dilakukan pengujian akurasi dengan memasukkan satu persatu variabel, yang kemudian pada contoh dibawah ini ditampilkan ketika variabel sudah mencapai 10.

```

[100] In: >>> model_classification(x, y, model=x)
sfs = SequentialFeatureSelector(k=10, n_features_to_select=10)
sfs.fit(X, y)
feature_selection_accuracy(objective=model_classification(x, y, model=x), #function 'model', n_features_to_select=10)
sfs.get_support()
sfs.transform(X).shape

(83, 10)

```

Gambar 4. 5. Perhitungan menggunakan 10 variabel

Yang dimana variabel yang terpilih dapat dilihat pada Gambar 4.6

```

[100] In: >>> print("features selected by forward sequential selection: " + sfs.get_feature_names_out())

features selected by forward sequential selection: ['NILAI SEMESTER' 'PERAN ORANGTUA' 'PELAKSANAAN EVALUASI'
'PEMBERIAN MOTIVASI' 'TINGKAT' 'BTQ' 'KEAKTIFAN' 'MENYAT' 'MOTIVASI'
'UMUR']

```

Gambar 4. 6. 10 variabel terpilih

Yang dimana hasil contoh dataset dapat dilihat pada gambar 4.7

	NILAI SEMESTER	PERAN ORANGTUA	PELAKSANAAN EVALUASI	PUNYAIAN MOTIVASI	TINGKAT	BTQ	KEAKTIFAN	PENYAT	MOTIVASI	UMUR	STATUS KELUARGA
0	90	87	83	75	90	88	85	75	85	12	1
1	90	25	83	72	90	88	81	87	82	12	1
2	75	87	86	75	90	88	85	86	86	12	1
3	70	88	83	72	90	88	83	84	85	12	1
4	85	85	85	75	90	25	75	74	86	12	0

Gambar 4. 7. contoh dataset ketika variabel berjumlah 10

Ketika variabel berjumlah 10, menghasilkan nilai akurasi sebesar

```
for i in range(2, 9): #{initial,final but not included,gap}
    knn = KNeighborsClassifier(n_neighbors=i)
    knn.fit(X_train, y_train)
    print(f"k-value = , {i}")
    print(knn.score(X_test, y_test))
    print(f"-----")

k-value = , 2
0.7692307692307693
-----
k-value = , 3
0.8461538461538461
-----
k-value = , 4
0.8461538461538461
-----
k-value = , 5
0.7692307692307693
-----
k-value = , 6
0.8461538461538461
-----
k-value = , 7
0.8461538461538461
-----
k-value = , 8
0.7692307692307693
-----
```

Gambar 4. 8. akurasi KNN ketika variabel berjumlah 10

Kemudian dimasukkan variabel baru, dalam hal ini karena telah memasukkan 10 variabel, maka selanjutnya peneliti memasukkan 11 variabel, kemudian dilakukan pengukuran akurasi menggunakan KNN yang menghasilkan akurasi.

```

✓ [115] for i in range(2, 9): #(initial,final but not included,gap)
ed     knn = KNeighborsClassifier(n_neighbors=i)
        knn.fit(X_train, y_train)
        print(f"k-value = , {i}")
        print(knn.score(X_test, y_test))
        print(f"-----")

k-value = , 2
0.7692307692307693
-----
k-value = , 3
0.7692307692307693
-----
k-value = , 4
0.8461538461538461
-----
k-value = , 5
0.7692307692307693
-----
k-value = , 6
0.8461538461538461
-----
k-value = , 7
0.7692307692307693
-----
k-value = , 8
0.7692307692307693
-----

```

Gambar 4. 9. akurasi KNN ketika variabel berjumlah 11

Dapat terlihat pada penelitian menggunakan 11 variabel, yang dimana tidak memberi dampak yang signifikan sehingga variabel yang terpilih adalah 10.

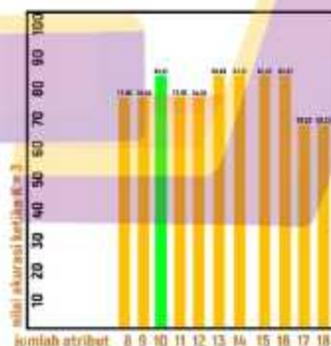
2. Setelah melakukan pengujian pada setiap kombinasi fitur, kita memilih model KNN dengan akurasi tertinggi sebagai model terbaik. Dalam contoh ini, model KNN dengan 10 variabel serta  $K = 3$  memiliki akurasi tertinggi, sehingga dipilih sebagai model terbaik untuk memprediksi target variable.

berikut hasil perhitungan target hafalan siswa menggunakan KNN dan forward selection, pada proses ini nilai  $K$  mengalami perubahan dari 2 hingga 8 yang hasilnya ditampilkan pada tabel 4.10.

Tabel 4. 10. Hasil Perhitungan KNN + Forward Selection

Nilai K	Akurasi
2	76.92%
3	84.61%
4	84.61%
5	76.92%
6	84.61%
7	84.61%
8	76.92%

Pada tabel di atas, dapat terlihat bahwa nilai akurasi terbaik adalah ketika K bernilai 3 yaitu mencapai 84.61%, sedangkan setelah nilai K ditingkatkan tidak terjadi peningkatan, sehingga didapatkan nilai terbaik adalah ketika K bernilai 3, kemudian dilakukan perbandingan dengan percobaan lain dengan jumlah atribut yang berbeda. Adapun grafik perbandingan percobaan pencarian seleksi atribut terbaik dapat dilihat pada gambar 4.11.



Gambar 4. 10. grafik perbandingan akurasi ketika K bernilai 3

Pada gambar di atas dapat dilihat bahwa ketika atribut berjumlah 10 dan nilai  $K=3$  menghasilkan akurasi tertinggi yaitu 84.61% jika dibandingkan dengan percobaan lainnya, alasan mengapa dipilih ketika atribut bernilai 10 meski ketika atribut berjumlah 13,14,15, dan 16 juga menghasilkan akurasi yang sama yaitu 84.61%, karena dalam hal ini peneliti sedang menyeleksi atribut sehingga semakin sedikit atribut yang ada maka akan semakin memudahkan proses pengelompokan pada program karantina Al-Qur'an.

3. Mengukur model menggunakan Metrik akurasi digunakan untuk menilai performa model machine learning dalam mengelompokkan sesuatu ke dalam kelas yang benar.

```

✓ [190] print(classification_report(y_test,y_pred))
ou

```

	precision	recall	f1-score	support
0	0.71	0.83	0.77	6
1	0.83	0.71	0.77	7
accuracy			0.77	13
macro avg	0.77	0.77	0.77	13
weighted avg	0.78	0.77	0.77	13

Gambar 4. 11. Metrik evaluasi KNN + Forward Selection

Pada hasil Metrik evaluasi KNN + FS pada Gambar 4. 23. Model memprediksi kelas 0 dengan precision 0.71, recall 0.83, dan f1-score 0.77. Artinya, model memprediksi kelas 0 dengan benar 71% dari waktu dan dapat menemukan 83% kejadian kelas 0 yang sebenarnya ada. Model juga memprediksi kelas 1 dengan precision 0.83, recall 0.71, dan f1-score 0.77. Artinya, model memprediksi kelas 1

dengan benar 83% dari waktu dan dapat menemukan 71% kejadian kelas 1 yang sebenarnya ada.

Berikut Confusion Matrix algoritma KNN pada saat K bernilai 3

Tabel 4. 11. Confusion Matrix KNN

	True LULUS	True TIDAK LULUS	Class precision
pred LULUS	5	1	71%
Pred. TIDAK LULUS	2	5	83%
Class recall	83%	71%	

Keterangan tabel 4.11. adalah:

- Banyaknya data yang seharusnya LULUS kemudian diprediksi LULUS sebanyak 5.
- Banyaknya data yang seharusnya TIDAK LULUS kemudian diprediksi TIDAK LULUS sebanyak 5.
- Banyaknya data yang seharusnya TIDAK LULUS kemudian diprediksi LULUS sebanyak 2.
- Banyaknya data yang seharusnya LULUS kemudian diprediksi TIDAK LULUS sebanyak 1.

Evaluasi pada model mengaplikasikan nilai akurasi dan classification error.

Akurasi dan classification error dari model yaitu :

$$\text{Akurasi} = \frac{5+5}{5+1+2+5} = 85,00\%$$

$$Error = \frac{2+1}{5+1+2+5} = 15,00\%$$

#### 4.2.3. Penerapan Algoritma PSO pada KNN

Pada metode KNN + PSO dihadirkan tabel yang memberikan informasi pemberian nilai awal serta simbol yang digunakan pada prosesnya.

Tabel 4. 12. Simbol-simbol yang digunakan dalam proses perhitungan metode KNN + PSO

Simbol	Keterangan	Nilai yang diterapkan
w	Bobot inersia 0-1	0,9
c1, c2	Learning rates	2,2
r1, r2	Bilangan random yang - berdistribusi dalam interval 0 - 1	1,0
f	Fitness	
i	Iterasi	100
x	Partikel	
j	Nomer partikel	
n	Jumlah partikel	20
v	Kecepatan	

Berikut contoh penerapannya pada model yang digunakan dalam proses seleksi fitur (feature selection) dengan menggunakan metode particle swarm optimization. Dengan menggunakan fungsi ini, maka dapat dilakukan optimasi terhadap seluruh fitur yang ada dalam data latih, sehingga dapat ditentukan fitur-fitur mana saja yang paling bermanfaat dalam menentukan target dari data latih.. Fungsi ini dapat dilihat pada Gambar 4.13.

```

def jfs(xtrain, ytrain, opts):
    # Parameters
    ub = 1
    lb = 0
    thres = 0.5
    w = 0.9 # inertia weight
    c1 = 2 # acceleration factor
    c2 = 2 # acceleration factor

    N = opts['N']
    max_iter = opts['T']
    if 'w' in opts:
        w = opts['w']
    if 'c1' in opts:
        c1 = opts['c1']
    if 'c2' in opts:
        c2 = opts['c2']

    # Dimension
    dim = np.size(xtrain, 1)
    if np.size(lb) == 1:
        ub = ub * np.ones([1, dim], dtype='float')
        lb = lb * np.ones([1, dim], dtype='float')

```

Gambar 4. 12. Menerapkan algoritma optimisasi partikel untuk memilih subset fitur terbaik untuk regresi linier

Selanjutnya sebuah variabel bernama `opts` yang akan digunakan sebagai parameter-parameter yang akan digunakan dalam proses seleksi fitur menggunakan metode PSO. Seperti pada Gambar 4.14.

```

c1 = 2 # cognitive factor
c2 = 2 # social factor
w = 0.9 # inertia weight
k = 5 # k-value in KNN
N = 20 # number of population
T = 100 # maximum number of iterations
opts = {'k':k, 'fold':fold, 'N':N, 'T':T, 'w':w, 'c1':c1, 'c2':c2}

```

Gambar 4. 13. Mengatur parameter untuk algoritma optimisasi partikel untuk pemilihan fitur dalam regresi linier

Berikut langkah-langkah perhitungan manual metode KNN + PSO

1. Menentukan jumlah partikel ( $n$ ) serta memberikan nilai acak pada posisi partikel atau nilai  $K$  nya,  $n=20$ .  $X$  merupakan partikel ke berapa, Pada tabel 4.13 terdapat  $n$  dan  $K$  yang telah ditentukan.

Tabel 4. 13. Tabel jumlah partikel serta nilainya dimasing-masing partikel.

No.	Partikel	Nilai Posisi Partikel (angka acak)
1	$X_1(0)$	35
2	$X_2(0)$	5
3	$X_3(0)$	21
4	$X_4(0)$	15
5	$X_5(0)$	10
6	$X_6(0)$	7
...	...	...
...	...	...
20	$X_{20}(0)$	1

2. Selanjutnya mencari error rate pada setiap partikel. dengan rumus Error rate = (jumlah data yang salah diklasifikasikan / total jumlah data) x 100%. Contoh perhitungannya adalah:

$K=35$ ; jumlah data yang salah diklasifikasikan = 20; total jumlah data=63

Error rate=  $(20/63)100=31$ , sehingga ketika  $K$  bernilai 35 eror ratenya adalah 31

Tabel 4. 14. Nilai error rate di setiap partikel

No.	K	Error Rate %
1	35	31
2	23	25
3	21	19
4	15	17
5	11	15
6	7	30
...	...	...
...	...	...
20	1	29

3. Kemudian mengevaluasi nilai fitness fungsi tujuan pada tiap partikel  $X_j(i)$  untuk  $j=1,2,3,4, \dots, 20$ , pada iterasi ke-0 ini error masih 100%. Berikut nilai fungsi disetiap partikel, Fungsi tujuan atau nilai fitness disimbolkan dengan F.

Tabel 4. 15. Hasil nilai fitness evaluasi disetiap partikel pada iterasi ke-0

No.	Partikel	Nilai posisi partikel
1	$Fx_1(0)$	100
2	$Fx_2(0)$	100
3	$Fx_3(0)$	100
4	$Fx_4(0)$	100
5	$Fx_5(0)$	100
6	$Fx_6(0)$	100
...	...	...
20	$Fx_{20}(0)$	100

4. Tentukan  $P_{best}$  dan  $G_{best}$ , untuk  $P_{best}$  digunakan rumus jika  $f_{xj}(i) < f_{xj}(i-1)$  maka  $P_{best} = x_{1j}(i)$ . Sedangkan nilai  $G_{best}$  merupakan nilai minimum fungsi tujuan diantara semua partikel untuk semua iterasi sebelumnya  $f_{xj}(i)$ . Pada iterasi ke-0 belum ada data yang bisa dihadirkan.

Tabel 4. 16. Hasil  $P_{best}$  dan  $G_{best}$  pada iterasi ke-0

No.	Partikel	Nilai posisi partikel	Nilai $G_{best}$
1	$F_{x1}(0)$	-	-
2	$F_{x2}(0)$	-	-
3	$F_{x3}(0)$	-	-
4	$F_{x4}(0)$	-	-
5	$F_{x5}(0)$	-	-
6	$F_{x6}(0)$	-	-
...	...	...	-
...	...	...	-
20	$F_{x20}(0)$	-	-

5. Menentukan kecepatan awal ( $V$ ) partikel pada iterasi ke-0, yaitu  $V_1(0)=V_2(0)=V_3(0)=V_4(0)=V_5(0)=V_6(0) \dots V_{24}(0)=V_{20}(0)=0$ . Pada iterasi ke-0 data masih 0.

Tabel 4. 17. Hasil kecepatan awal partikel pada iterasi ke-0

No.	Partikel	Nilai posisi partikel
1	$V_1(0)$	0
2	$V_2(0)$	0

Tabel 4. 17. (Lanjutan)

No.	Partikel	Nilai posisi partikel
3	$V_3(0)$	0
4	$V_4(0)$	0
5	$V_5(0)$	0
6	$V_6(0)$	0
...	...	...
...	...	...
20	$V_{20}(0)$	0

6. Kemudian beranjak ke langkah awal pada iterasi selanjutnya yaitu iterasi ke-1. Untuk mendapatkan nilai posisi partikel pada iterasi ke-1 digunakan  $X_j(i) = X_j(i-1) + V_j(i)$ . Partikel ke-1 iterasi ke-1 disimbolkan dengan  $X_1(1)$ .

Contoh mencari nilai posisi partikel ke-1 iterasi ke-1 yaitu:  $X_1(1) = 35 + 0 = 35$ .

Nilai 35 merupakan nilai posisi partikel ke-1 iterasi ke-0  $X_j(i-1)$  sedangkan 0 adalah kecepatan partikel ke-1 iterasi ke-0  $V_j(i)$ . Kemudian dihitung sampai partikel ke-20.

Tabel 4. 18. Hasil nilai posisi partikel pada iterasi ke-1

No.	Partikel	Nilai Posisi Partikel
1	$X_1(1)$	35
2	$X_2(1)$	5
3	$X_3(1)$	21

Tabel 4. 18. (Lanjutan)

No.	Partikel	Nilai Posisi Partikel
4	$X_4(1)$	15
5	$X_5(1)$	12
6	$X_6(1)$	7
...	...	...
...	...	...
20	$X_{20}(1)$	1

7. Dalam studi ini, digunakan sebanyak 63 baris data untuk dicari akurasiya menggunakan metode KNN. Sebagai contoh, mencari nilai posisi partikel pada fitness partikel ke-1. Data tersebut dimasukkan dan K-nya ditentukan, seperti  $X_j(i)$  dengan nilai K 35, yang didapat dari  $X_1(1)$  partikel ke-1 iterasi ke-1. Setelah akurasiya dihitung, diperoleh hasil pencarian fitness pada partikel ke-1 iterasi ke-1. Dalam hal ini, akurasi 100% dikurangi dengan hasil akurasi perhitungan sesuai dengan jumlah K, misalnya mencari fitness partikel ke-1 sebesar 69%, maka  $100-69=31$ . Oleh karena itu, nilai partikel pada fitness partikel ke-1 iterasi ke-1 adalah 31%. Kemudian, perhitungan dilakukan hingga mencapai nilai fitness pada iterasi ke-20. Berikut ini adalah hasil posisi nilai partikel pada semua fitness partikel.

Tabel 4. 19. Hasil fitness nilai evaluasi setiap partikel pada iterasi ke-1

No.	Partikel	Nilai posisi partikel
1	$Fx_1(1)$	31
2	$Fx_2(1)$	25

Tabel 4. 19. (Lanjutan)

No.	Partikel	Nilai posisi partikel
3	$Fx_3(1)$	19
4	$Fx_4(1)$	17
5	$Fx_5(1)$	15
6	$Fx_6(1)$	30
...	...	...
...	...	...
20	$Fx_{20}(1)$	29

8. Mencari nilai  $P_{best}$  dan  $G_{best}$ . Untuk mendapatkan nilai posisi partikel pada  $P_{best}$  digunakan rumus, jika  $fx_j(i) < fx_j(i-1)$  maka  $P_{best} = x_1(i)$ .

Sebagai contoh perhitungan pada  $P_{best}$   $x_1 =$  jika  $11 < 100$ , maka 35.

Perhitungan diteruskan hingga  $P_{best}$  ke-20. Untuk nilai  $G_{best}$  ditentukan dengan cara mencari fungsi tujuan terkecilnya disimbolkan dengan  $Fx_j(i)$ . Berikut ditampilkan seluruh nilai masing-masing partikel  $P_{best}$  dan nilai  $G_{best}$ .

Tabel 4. 20. Hasil  $P_{best}$  dan nilai  $G_{best}$  pada iterasi ke-1

No.	Partikel	Nilai posisi partikel	Nilai $G_{best}$
1	$P_{best} X_1$	35	
2	$P_{best} X_2$	25	
3	$P_{best} X_3$	21	
4	$P_{best} X_4$	17	

Tabel 4. 20. (Lanjutan)

No.	Partikel	Nilai posisi partikel	Nilai Gbest
5	$P_{best} X_5$	15	6
6	$P_{best} X_6$	30	
...	...	...	
...	...	...	
20	$P_{best} X_{20}$	29	

Gbest terbaik adalah ketika  $F_{X_5} = 10$  posisi partikel ( $P_{best}$ ) dan (fitness atau fungsi tujuan minimum adalah 6)

9. Dilakukan perhitungan kecepatan  $V_j(i)$  untuk memperoleh nilai kecepatan pada setiap partikel. Untuk itu, digunakan perhitungan kecepatan pada partikel ke-1 pada iterasi ke-1.

$V_j(i) = w \times V \text{ lama} + c_1 \times r_1 \times (p \text{ best} - \text{posisi partikel lama}) + c_2 \times r_2 \times (g \text{ best} - \text{posisi partikel lama})$ .  
 $V_j(i) = 0.9 \times 0 + 2 \times 1 \times (35-35) + 2 \times 0 \times (6-35) = 0$ .

Kemudian teruskan langkah di atas untuk semua partikel.

Tabel 4. 21. Hasil kecepatan partikel  $V_j(i)$  pada iterasi ke-1

No.	Partikel	Nilai posisi partikel
1	$V_1(1)$	0
2	$V_2(1)$	0
...	...	...
20	$V_{20}(1)$	0

10. Pada perhitungan Pbest dan Gbest pada iterasi 1 belum konvergen atau nilai antara Pbest dan Gbest belum sama, maka akan dilanjutkan ke iterasi ke-2. berikut hasil dari iterasi yang dilakukan dapat dilihat pada Gambar 4. 16.

```

# perform feature selection
start_time = time.time()
fmdl = jfs(X_train, y_train, opts)
print("Run Time ---%s seconds ---" % (time.time() - start_time))

sf = fmdl['sf']

# model with selected features
num_train = np.size(xtrain, 0)
num_valid = np.size(xtest, 0)
X_train = xtrain[:, sf]
y_train = ytrain.reshape(num_train) # Solve bug
X_valid = xtest[:, sf]
y_valid = ytest.reshape(num_valid) # solve bug

mdl = linearRegression()
mdl.fit(X_train, y_train)

# accuracy
y_pred = mdl.predict(X_valid)
RMSE = mean_squared_error(y_valid, y_pred, squared=False)
print("RMSE:", RMSE)

# number of selected features
num_feat = fmdl['nf']
print("Feature Size:", num_feat)

# plot convergence
curve = fmdl['c']
curve = curve.reshape(np.size(curve,1))
x = np.arange(0, opts['i'], 1.0) + 1.0

fig, ax = plt.subplots()
ax.plot(x, curve, 'o-')
ax.set_xlabel('Number of Iterations')
ax.set_ylabel('Fitness')
ax.set_title('PSO')
ax.grid()
plt.show()

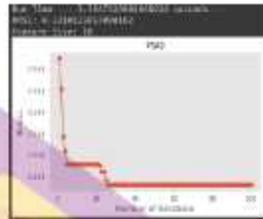
```

Gambar 4. 14. source code proses pada PSO

```

Iteration: 1
Best (PSO): 0.34494742929753325
Iteration: 2
Best (PSO): 0.3421101923336219
Iteration: 3
Best (PSO): 0.3377153060989836
Iteration: 4
Best (PSO): 0.3363444686778948
Iteration: 5
Best (PSO): 0.33516136521755024
Iteration: 6
Best (PSO): 0.33516136521755024
Iteration: 7
Best (PSO): 0.33516136521755024
Iteration: 8
Best (PSO): 0.33516136521755024
Iteration: 9
Best (PSO): 0.33516136521755024
Iteration: 10
Best (PSO): 0.33516136521755024

```



Gambar 4. 15. PSO Iteration

Dalam setiap iterasi, PSO akan menghitung nilai fitness dari setiap partikel, dan partikel dengan nilai fitness yang lebih tinggi akan dijadikan sebagai pemimpin lokal (local best) dan pemimpin global (global best) pada populasi. Partikel-partikel pada populasi kemudian akan diperbarui posisinya berdasarkan percepatan yang dihitung dari perbedaan posisi partikel dengan pemimpin lokal dan global.

Proses ini dilakukan secara berulang-ulang hingga mencapai kondisi berhenti yang telah ditentukan sebelumnya, seperti mencapai jumlah iterasi maksimum atau mencapai nilai fitness tertentu. Setelah proses seleksi atribut selesai, akan didapatkan himpunan fitur yang dianggap paling optimal untuk digunakan pada model pembelajaran. Jumlah fitur terpilih juga dicetak dan berjumlah 10.

Dengan hasil fitur terseleksi adalah sebagai berikut

```
[ ] feature_selection = []

items= fmdl['sf']
for index, item in enumerate(items):
    print(f"{item} - {df2.columns[item]}")
    feature_selection.append(df2.columns[item])

2 = PERAN ORANGTUA
4 = PELAKSANAAN EVALUASI
5 = POLA KOMUNIKASI
7 = TINGKAT
8 = BTQ
9 = KEAKTIFAN
10 = KEPERIBADIAN
12 = MINAT
16 = UMUR
17 = JENIS KELAMIN
```

Gambar 4. 16. Hasil selection fitur PSO

Setelah proses optimasi selesai, fitur-fitur yang terpilih dapat ditentukan berdasarkan posisi partikel global-best. Fitur-fitur dengan nilai bobot yang besar pada posisi tersebut dianggap sebagai fitur terpilih. Berikut contoh dataset setelah dilakukan feature seleksi menggunakan PSO

```
1. #load the dataset
df = pd.read_csv('data.csv')
df = pd.DataFrame(df, columns = feature_selection + ['STATUS KELUHAN'])

#print the first 5 rows of the dataframe
df.head()
```

	PERAN ORANGTUA	PELAKSANAAN EVALUASI	POLA KOMUNIKASI	TINGKAT	BTQ	KEAKTIFAN	KEPERIBADIAN	MINAT	UMUR	JENIS KELAMIN	STATUS KELUHAN
0	87	85	85	90	88	85	86	78	12	2	1
1	84	83	84	90	88	85	86	87	11	2	1
2	87	84	84	90	88	85	86	88	12	2	1
3	80	82	81	90	80	81	86	86	12	1	1
4	81	83	70	80	25	75	75	78	12	1	0

Gambar 4. 17. Contoh tampilan dataset

```

for i in range(2, 9): #(initial,final but not included,step)
    knn = KNeighborsClassifier(n_neighbors=i)
    knn.fit(X_train, y_train)
    print(f"K-value = , {i}")
    print(knn.score(X_test, y_test))
    print(f"-----")

```

```

K-value = , 2
0.7692307692307693
-----

```

```

K-value = , 3
0.7692307692307693
-----

```

```

K-value = , 4
0.7692307692307693
-----

```

```

K-value = , 5
0.6923076923076923
-----

```

```

K-value = , 6
0.8461538461538461
-----

```

```

K-value = , 7
0.8461538461538461
-----

```

```

K-value = , 8
0.8461538461538461
-----

```

Gambar 4. 18. Akurasi K value KNN berdasarkan PSO

Setelah diketahui variabel yang paling mempengaruhi maka diteruskan untuk dilakukan perhitungan menggunakan KNN, pada proses perhitungan menggunakan KNN nilai K pada proses ini diubah sehingga dapat dilakukan perbandingan nilai akurasi, adapun hasilnya dapat dilihat pada tabel 4.22.

Tabel 4. 22. Hasil Perhitungan KNN + PSO

Nilai K	Akurasi
2	0.7692
3	0.7692
4	0.7692
5	0.6923

Tabel 4. 22. (Lanjutan)

Nilai K	Akurasi
6	0.8461
7	0.8461
8	0.8461

Pada tabel di atas, dapat terlihat bahwa nilai akurasi terbaik adalah ketika K bernilai 6 yaitu mencapai 84.61%, sedangkan setelah nilai K ditingkatkan tidak terjadi peningkatan, sehingga didapatkan nilai terbaik adalah ketika K bernilai 6. Pada Gambar 4.20. Precision, recall, f1-score, dan lainnya disebut dengan metrik evaluasi model machine learning atau metrik akurasi. Metrik akurasi digunakan untuk menilai performa model machine learning dalam mengelompokkan sesuatu ke dalam kelas yang benar.

```

f | .print(classification_report(y_test,y_pred))

```

	precision	recall	f1-score	support
0	0.75	1.00	0.86	6
1	1.00	0.71	0.83	7
accuracy			0.85	13
macro avg	0.88	0.86	0.85	13
weighted avg	0.88	0.85	0.84	13

Gambar 4. 19. Metrik evaluasi KNN

Berdasarkan hasil metrik evaluasi KNN + PSO pada Gambar 4.20. dapat disimpulkan beberapa hal diantaranya:

Berikut Confusion Matrix algoritma KNN pada saat K bernilai 6

Tabel 4. 23. Confussion Matrix KNN

	True LULUS	True TIDAK LULUS	Class precision
pred LULUS	6	0	75%
Pred. TIDAK LULUS	2	5	100%
Class recall	100%	71%	

Keterangan tabel 4.23. adalah:

- Banyaknya data yang seharusnya LULUS kemudian diprediksi LULUS sebanyak 6.
- Banyaknya data yang seharusnya TIDAK LULUS kemudian diprediksi TIDAK LULUS sebanyak 5.
- Banyaknya data yang seharusnya TIDAK LULUS kemudian diprediksi LULUS sebanyak 2.
- Banyaknya data yang seharusnya LULUS kemudian diprediksi TIDAK LULUS sebanyak 0.

Evaluasi pada model mengaplikasikan nilai akurasi dan classification error.

Akurasi dan classification error dari model yaitu :

$$\text{Akurasi} = \frac{6+5}{6+0+2+5} = 85,00\%$$

$$\text{Error} = \frac{2+0}{6+0+2+5} = 15,00\%$$

#### 4.3. Hasil Penelitian

Hasil yang didapatkan setelah melakukan perhitungan menggunakan algoritma KNN dengan menggunakan fitur terbaik yang didapatkan dari algoritma Forward

Selection dan PSO, menghasilkan tingkat akurasi masing-masing sebagai berikut dapat dilihat pada Tabel

Tabel 4. 24. perbandingan akurasi

K	KNN	KNN + Forward Selection	KNN + PSO
2	61.53%	76.92%	0.7692
3	69.23%	84.61%	0.7692
4	53.84%	84.61%	0.7692
5	69.23%	76.92%	0.6923
6	69.23%	84.61%	0.8461
7	69.23%	84.61%	0.8461
8	69.23%	76.92%	0.8461

Berdasarkan nilai akurasi di atas, dapat diketahui bahwa penambahan fitur dengan menggunakan algoritma Forward Selection dan PSO memberikan hasil yang lebih maksimal jika dibandingkan hanya menerapkan algoritma KNN saja yang hanya memberikan akurasi 69% ketika K=3. Pada nilai K tertentu, penambahan fitur dengan menggunakan algoritma Forward Selection mendapatkan akurasi tertinggi ketika K=3 dengan akurasi senilai 84.61% kemudian PSO memberikan hasil akurasi tertinggi ketika K=6 dengan akurasi senilai 84.61%, perbedaan nilai K terbaik terjadi dikarenakan perbedaan atribut yang terpilih pada masing-masing percobaan, yang dimana atribut pada percobaan algoritma FS berbeda dengan atribut terpilih pada PSO, hal ini menjadikan perbedaan akurasi terbaik dengan nilai K berbeda.

Maka didapatkan faktor yang sangat berpengaruh dalam perhitungan akurasi target hafalan Al-Qur'an pada SMPIT MIS ada pada tabel 4.25.

Tabel 4. 25. Tabel Perbandingan fitur terbaik dalam mempengaruhi akurasi

KNN + Forward Selection	KNN + PSO
MINAT PELAKSANAAN EVALUASI UMUR PERAN ORANGTUA TINGKAT BTQ KEAKTIFAN NILAI SEMESTER MOTIVASI PEMBERIAN MOTIVASI	MINAT PELAKSANAAN EVALUASI UMUR PERAN ORANGTUA TINGKAT BTQ KEAKTIFAN POLA KOMUNIKASI KEPRIBADIAN JENIS KELAMIN

Sehingga faktor yang dianggap tidak begitu berpengaruh dalam perhitungan akurasi target hafalan siswa SMPIT MIS dapat dilihat pada tabel 4.26.

Tabel 4. 26. Faktor yang tidak mempengaruhi target hafalan siswa

No.	Faktor yang tidak begitu mempengaruhi akurasi
1	Reward
2	Kerjasama
3	Perubahan Lingkungan
4	Disiplin waktu
5	Adaptasi

Berdasarkan hasil seleksi, maka telah diketahui 18 dari faktor yang ada, 13 diantaranya dinilai mempengaruhi akurasi secara signifikan sehingga 5 faktor lainnya dianggap kurang/tidak terlalu mempengaruhi akurasi.

Setelah diketahui faktor yang paling mempengaruhi target hafalan Al-Qur'an, maka nantinya faktor tersebut akan digunakan pada program karantina Al-Qur'an untuk mengelompokkan siswa, sehingga pembagian siswa menjadi seimbang dengan menempatkan siswa yang diprediksi lulus dan siswa yang diprediksi tidak lulus dalam satu kelompok yang sama, yang nantinya guru pengampu bisa lebih maksimal dalam membimbing dan fokus dalam membagi

perhatiannya dan setiap siswa diharapkan dapat menyelesaikan permasalahan yang ada.

Pada penelitian sebelumnya (Ting Zou, DKK, 2021) yang menilai bahwa kerjasama dapat lebih memudahkan anak dalam mengingat serta mengurangi bahkan menghilangkan kecurangan, pada penelitian ini yang dimana membahas mengenai hafalan Al-Qur'an dinilai tidak sama pengaruhnya dikarenakan pengaruhnya terhadap akurasi dinilai memiliki bobot yang kurang tinggi. Kemudian penelitian mengenai faktor usia (Ahsin Al-Hafidz, 2005) yang dimana menilai faktor usia dapat mempengaruhi kelancaran dalam mempengaruhi hafalan. Sedangkan pada penelitian ini dianggap sama pengaruhnya terhadap hasil akurasi. Penelitian ini memiliki hasil yang sama dengan penelitian yang dilakukan oleh (Yaumiyatul Furqoni dan Lilik Mufidah, 2019) yang dimana menganggap bahwa faktor disiplin dalam kedatangan jam sekolah dan jam pulang tidak mempengaruhi hafalan siswa, sedangkan pada penelitian ini faktor tersebut yang diujikan pada siswa SMPIT MIS terbukti sama dengan penelitian sebelumnya yaitu tidak mempengaruhi target hafalan mereka.

Sehingga kedepannya pihak SMPIT MIS bisa menerapkan dan memerperhitungkan variabel yang dianggap mempengaruhi target hafalan siswa untuk mengelompokkan siswa ketika masa karantina Al-Qur'an dilaksanakan.

## BAB V

### PENUTUP

#### 5.1. Kesimpulan

Berdasarkan hasil pengujian yang telah dilakukan, maka dapat ditarik kesimpulan sebagai berikut:

1. Penerapan KNN dengan 18 variabel menghasilkan tingkat akurasi tertinggi senilai 69% ketika nilai  $K=3$ .
2. Penerapan KNN yang telah dilakukan optimasi menggunakan FS menghasilkan atribut terpilih sebanyak 10 variabel dan didapatkan nilai akurasi terbaik adalah 84.61% ketika nilai  $K=3$ .
3. Penerapan KNN yang telah dilakukan optimasi menggunakan PSO menghasilkan variabel terpilih sebanyak Sehingga didapatkan nilai akurasi terbaik adalah 84.61% ketika nilai  $K=6$ .
4. Setelah dilakukan perbandingan antara KNN, KNN dan FS, kemudian KNN dan PSO, maka didapatkan akurasi terbaik dihasilkan ketika KNN telah dilakukan optimasi baik itu ketika menggunakan FS maupun PSO yang dimana masing-masing algoritma memberikan dampak peningkatan yang sama yaitu 15%.
5. Variabel terbaik dalam mempengaruhi akurasi pada model KNN dan FS adalah Minat, Pelaksanaan Evaluasi, Umur, Peran Orangtua, Tingkat, Btq, Keaktifan, Nilai Semester, Motivasi, Pemberian Motivasi, sedangkan pada model KNN dan PSO adalah Minat, Pelaksanaan Evaluasi, Umur, Peran Orangtua, Tingkat, Btq, Keaktifan, Pola Komunikasi, Kepribadian, Jenis

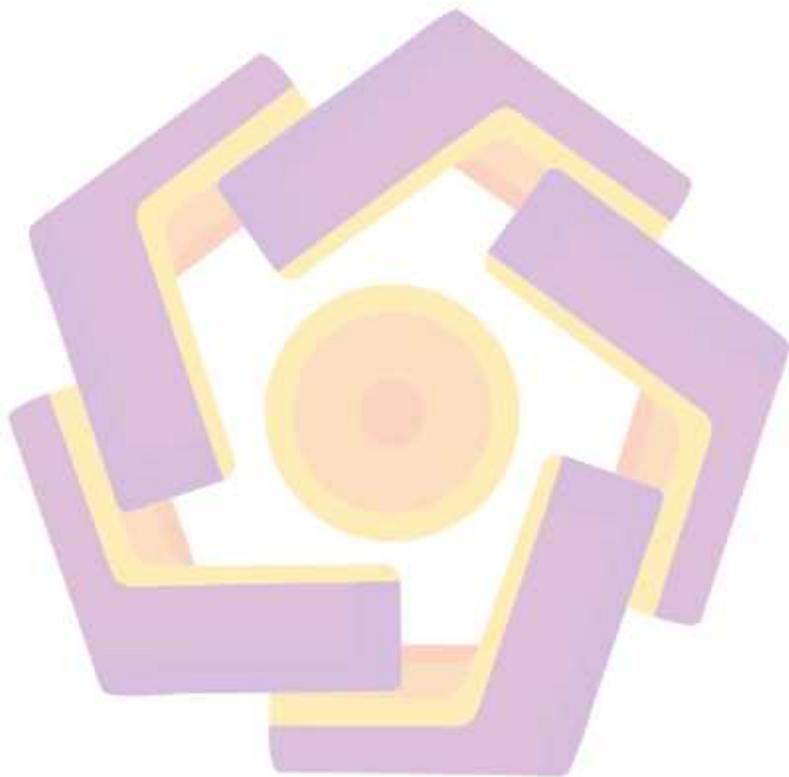
Kelamin.

6. Variabel yang dianggap kurang mempengaruhi akurasi dalam menentukan target hafalan siswa adalah Reward, Kerjasama, Perubahan Lingkungan, Disiplin waktu, Adaptasi.
7. Kelebihan dari metode Feature Selection Forward adalah bahwa metode ini menggunakan metode eliminasi langsung, yang berarti bahwa feature yang tidak penting akan langsung dikeluarkan dari model sehingga tidak perlu melakukan iterasi yang berulang-ulang. Sedangkan kelebihan dari metode PSO (Particle Swarm Optimization) adalah bahwa metode ini mudah diimplementasikan dan sederhana.
8. Kekurangan dari metode Forward Selection adalah bahwa metode ini bisa menjadi tidak efisien dalam kasus di mana ada banyak feature yang tersedia, karena metode ini harus mengevaluasi semua feature satu per satu. Selain itu, metode ini juga tidak selalu dapat menemukan solusi yang optimal. Sedangkan kekurangan dari metode PSO adalah bahwa metode ini dapat menjadi sangat sensitif terhadap pengaturan parameter yang digunakan, dan hasil dari metode ini juga bergantung pada asumsi yang dibuat tentang data yang tersedia.

## 5.2. Saran

1. Untuk penelitian selanjutnya dapat menggunakan algoritma lain dengan pertimbangan kecepatan algoritma agar proses yang dilakukan dapat lebih cepat.
2. Penelitian selanjutnya dapat menggunakan dataset yang lebih beragam.

3. Melakukan perbandingan dengan algoritma klaisifikasi lain yang telah sama-sama diterapkan Forward Selection dan PSO.
4. Meningkatkan tingkat akurasi pendeteksian target hafalan Al-Qur'an.



## DAFTAR PUSTAKA

### PUSTAKA BUKU

- Abbas, M. Ziyad. 1993. *Metode Praktis Menghafal al-Qur'an*, Jakarta: Firdaus. Alfaton, Sabit. 2019. *Teknik Menghafal Al-Qur'an*, Semarang: Mutiara Aksara.
- Abdurrah Nawabuddin dan Bambang Saiful Ma'arif, *Teknik Menghafal alQur'an (Kaifa Tahfiz al-Qur'an)*, Sinar Baru Algesindo, Bandung, 2005
- Ahsin Al-Hafidz, *Bimbingan Praktis Menghafal Al-Qur'an*, PT Bumi Aksara, Jakarta, 2005
- Abdul Khaliq, Syaikh Abdurrahman*. 1996. *Metode dan Strategi Dakwah Islam*. Jakarta: Pustaka Al-Kautsar. Addaruqutni, Imam. 2013.

### PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

- A. A. Aburomman and M. Bin Ibne Reaz, "A novel SVM-kNN-PSO ensemble method for intrusion detection system," *Appl. Soft Comput. J.*, vol. 38, pp. 360–372, 2016, doi: 10.1016/j.asoc.2015.10.011.
- N. Musyaffa and B. Rifai, "Model Support Vector Machine Berbasis Particle Swarm," *Jtik*, vol. 3, no. 2, pp. 189–194, 2018.
- H. Hendra, M. A. Azis, and S. Suhardjono, "Analisis Prediksi Kelulusan Mahasiswa Menggunakan Decision Tree Berbasis Particle Swarm Optimization," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 1, pp. 102–107, 2020, doi: 10.32736/sisfokom.v9i1.756.
- H. Muhabatin, C. Prabowo, I. Ali, C. L. Rohmat, and D. R. Amalia, "Klasifikasi Berita Hoax Menggunakan Algoritma Naïve Bayes Berbasis PSO," *INFORMATICS Educ. Prof. J. Informatics*, vol. 5, no. 2, pp. 156–165, 2021, [Online]. Available: <http://ejournal-binainsani.ac.id/index.php/ITBI/article/view/1531>.
- S. Budilaksono *et al.*, "Comparison of Data Mining Algorithm: PSO-KNN, PSO-RF, and PSO-DT to Measure Attack Detection Accuracy Levels on Intrusion Detection System," *J. Phys. Conf. Ser.*, vol. 1471, no. 1, 2020, doi: 10.1088/1742-6596/1471/1/012019.

- K. Sasirekha and K. Thangavel, "Optimization of K-nearest neighbor using particle swarm optimization for face recognition," *Neural Comput. Appl.*, vol. 31, no. 11, pp. 7935–7944, 2019, doi: 10.1007/s00521-018-3624-9.
- H. Meyer, C. Reudenbach, T. Hengl, M. Katurji, and T. Nauss, "Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation," *Environ. Model. Softw.*, vol. 101, pp. 1–9, 2018, doi: 10.1016/j.envsoft.2017.12.001.
- Fitriyah, D., 2008. Faktor yang mempengaruhi kecepatan menghafal Al-Qur'an antara santri mukim dan nonmukim di Pesantren Zaidatul Ma'arif Kauffman Parakan Temanggung. Skripsi. Semarang IAIN Walisongo.
- Gunawan, T.S., Ashraf, A., Riza, B.S., Haryanto, E.V., Rosnelly, R., Kartiwi, M., Janin, Z., 2020. Development of video-based emotion recognition using deep learning with Google Colab. *Telkomnika (Telecommunication Comput. Electron. Control.* 18, 2463–2471. <https://doi.org/10.12928/TELKOMNIKA.v18i5.16717>
- Oktarina, D., Yanti, Y.C.M., 2021. PENGARUH KOMUNIKASI INTERPERSONAL GURU DAN SANTRI TERHADAP MINAT HAFALAN AL-QUR'AN (Studi Pada Santri Putri Takhasus Pondok Pesantren Darul Huffaz Kecamatan Gedong Tataan Kabupaten Pesawaran Tahun 2020). *INTERCODE – J. Ilmu Komun.* Vol. 1 No. 1 (Maret, 2021), Hal. 49-62 ISSN 1, 35–48.
- Okwu, M.O., Tartibu, L.K., 2021. Particle Swarm Optimisation. *Stud. Comput. Intell.* 927, 5–13. [https://doi.org/10.1007/978-3-030-61111-8\\_2](https://doi.org/10.1007/978-3-030-61111-8_2)

Wijaya, C., Siregar, M.F.S., Ruslan, M., Holid, S., Roslaeni, R., 2021. Manajemen Pembelajaran Tahfiz dalam Peningkatan Minat Menghafal Alquran Siswa di Yayasan Tahfidzul Quran Al-Fawwaz Medan. *Edukasi Islam. J. Pendidik. Islam.* <https://doi.org/10.30868/ei.v10i01.1214>

