

**OPTIMASI PARAMETER K PADA ALGORITMA K-NN
UNTUK KLASIFIKASI JURUSAN PESERTA DIDIK BARU
DI MA DARUL HIKMAH MENGANTI**



SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Strata 1 (S.1) Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Nahdlatul Ulama Jepara

Oleh :

Maulana Ashar

NIM : 181240000759

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NAHDLATUL ULAMA JEPARA
2022**

PERSETUJUAN PEMBIMBING

Assalamu'allaikum Wr. Wb.

Setelah kami meneliti dan mengadakan perbaikan seperlunya, bersama ini saya kirim naskah skripsi Saudara:

Nama : Maulana Ashar
Nim : 181240000759
Program Studi : Teknik Informatika
Judul : Optimasi Parameter K Pada Algoritma KNN
Untuk Klasifikasi Jurusan Peserta Didik Baru Di
MA Darul Hikmah Menganti

Skripsi ini telah disetujui pembimbing dan siap untuk dipertahankan dihadapan Tim Penguji Program Sarjana Strata 1 (S1) Fakultas Sains dan Teknologi Universitas Islam Nahdlatul Ulama (Unisnu) Jepara.

Wassalamu'allaikum Wr. Wb.

Jepara, 4 Agustus 2022

Pembimbing I

Pembimbing II


Harminto Mulyo, S.Kom., M.Kom.

NIDN 0604028203

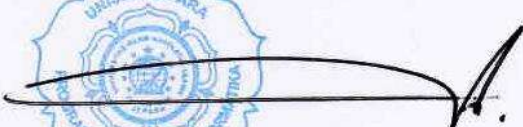

Nur Aeni Widiastuti, S.Pd., M.Kom.

NIDN 0602078702

Mengetahui,

Ketua Program studi Teknik Informatika




Gentur Wahyu Nyipto Wibowo, S.Kom, M.Kom.

NIDN 0623117902

PENGESAHAN

Skripsi dengan judul “Optimasi Parameter K Pada Algoritma K-NN Untuk Klasifikasi Jurusan Peserta Didik Baru Di MA Darul Hikmah Menganti” karya:

Nama : Maulana Ashar
Nim : 181240000759
Program Studi : Teknik Informatika

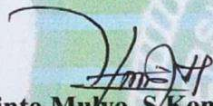
Telah diajukan dan dipertahankan dalam sidang oleh dewan penguji Fakultas Sains dan Teknologi Universitas Islam Nahdlatul Ulama (UNISNU) Jepara dan dinyatakan lulus pada tanggal : 22 Agustus 2022

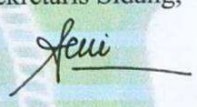
Selanjutnya dapat diterima sebagai syarat guna memperoleh gelar Sarjana Strata 1 (S1) Program Studi Teknik Informatika Fakultas Sains dan Teknologi UNISNU Jepara Tahun Akademik 2021/2022

Jepara, 22 Agustus 2022

Ketua Sidang,

Sekretaris Sidang,


Harminto Mulyo, S.Kom., M.Kom.

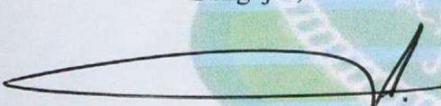

Nur Aeni Widiastuti, S.Pd., M.Kom.

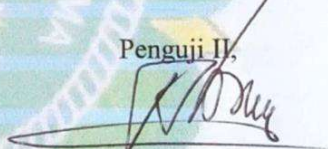
NIDN 0604028203

NIDN 0602078702

Penguji I,

Penguji II,


Gentur Wahyu Nyipto Wibowo, S.Kom, M.Kom.


Ir. Adi Sucipto, M.Kom

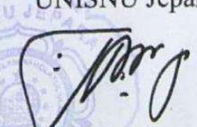
NIDN 0623117902

NIDN 0625056505

Mengetahui,

Dekan Fakultas Sain dan Teknologi

UNISNU Jepara


Dias Prihatmoko, S.T., M.Eng.

NIDN. 0612128302

PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Maulana Ashar
Nim : 181240000759
Program Studi : Teknik Informatika

Saya menyatakan dengan penuh kejujuran dan tanggung jawab, bahwa skripsi yang saya susun sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata 1 (S1) Universitas Islam Nahdlatul Ulama Jepara merupakan hasil karya saya sendiri yang jauh dari plagiarisme dan belum pernah diajukan sebagai pemenuhan persyaratan untuk memperoleh gelar Sarjana dari Perguruan Tinggi lain.

Adapun bagian-bagian tertentu dalam penulisan Skripsi yang saya kutip dari karya orang lain telah dituliskan sumbernya secara jelas dengan norma, kaidah, dan etika penulisan ilmiah.

Selanjutnya saya bersedia menerima sanksi dari Fakultas Sain dan Teknologi UNISNU Jepara apabila dikemudian hari ditemukan ketidakbenaran dari pernyataan ini.

Jepara, 4 Agustus 2022
Peneliti,



Maulana Ashar
NIM. 181240000759

ABSTRAK

Judul : Optimasi Parameter K Pada Algoritma K-NN Untuk
Klasifikasi Jurusan Peserta Didik Baru Di MA Darul Hikmah
Menganti
Penulis : Maulana Ashar
NIM : 181240000759
Program Studi : Teknik Informatika
Pembimbing I : Harminto Mulyo, S.Kom., M.Kom.
Pembimbing II : Nur Aeni Widiastuti, S.Pd., M.Kom.
Penguji I : Gentur Wahyu Nyipto Wibowo, S.Kom, M.Kom.
Penguji II : Ir. Adi Sucipto, M.Kom
Tanggal Ujian : 22 Agustus 2022

Penjurusan siswa merupakan suatu proses penempatan siswa ke dalam jurusan tertentu sesuai dengan minat dan kemampuan akademiknya sebagai upaya untuk lebih mempermudah siswa-siswi dalam proses pembelajaran. Madrasah Aliyah Darul Hikmah Menganti merupakan sekolah sederajat dengan SMA, yang memiliki dua jurusan yaitu IPA, dan IPS. Sulitnya dalam mengklasifikasikan jurusan peserta didik baru menjadi kendala bagi pihak sekolah. Karena proses penilaian kriteria yang dilakukan satu per satu. Dari permasalahan tersebut dilakukan penerapan metode *K-Nearest Neighbor* (K-NN) untuk mengklasifikasikan jurusan guna mempermudah dan meminimalisir kesalahan dalam proses penentuan jurusan siswa baru. Data yang awalnya berjumlah 638 record dan 31 atribut, setelah dilakukan preprocessing data yang digunakan berjumlah 635 record dengan 12 atribut yaitu nama, jenis kelamin, minat penjurusan, asal sekolah, anak ke, jumlah saudara, nilai matematika, nilai bahasa Inggris, nilai ipa, nilai bahasa Indonesia, nilai hasil tes, dan rekomendasi jurusan. Setelah dilakukan pengujian menggunakan *K-Fold Cross Validation* dan *Confusion Matrix* untuk evaluasi dan validasi hasil dengan perhitungan jarak *Euclidean Distance* didapatkan nilai k terbaik $k=3$ yang menghasilkan *accuracy*: 97.11%, *precision*: 96.82%, *recall*: 98.33%, dan AUC: 0.951.

Kata Kunci : Penjurusan siswa baru, Algoritma *K-Nearest Neighbor* (K-NN), *K-Fold Cross Validation*, *Confusion Matrix*

ABSTRACT

Title : Optimization of the K Parameter on the K-NN Algorithm for the Clasification Majoring Of New Students at MA Darul Hikmah Menganti
Author : Maulana Ashar
NIM : 181240000759
Study Program : Teknik Informatika
Supervisor I : Harminto Mulyo, S.Kom., M.Kom.
Supervisor II : Nur Aeni Widiastuti, S.Pd., M.Kom.
Tester I : Gentur Wahyu Nyipto Wibowo, S.Kom, M.Kom.
Tester II : Ir. Adi Sucipto, M.Kom
Test Date : 22 August 2022

Majoring students is a process of placing students into certain majors according to their interests and academic abilities in an effort to make it easier for students in the learning process. Madrasah Aliyah Darul Hikmah Menganti is a school equivalent to SMA, which has two majors, namely science and social studies. The difficulty in classifying the majors of new students is an obstacle for the school. Because the criteria assessment process is carried out one by one. From these problems, the K-Nearest Neighbor (K-NN) method was applied to classify majors in order to simplify and minimize errors in the process of determining new student majors. The data initially amounted to 638 records and 31 attributes, after preprocessing the data used amounted to 635 records with 12 attributes, namely name, gender, majoring interest, school origin, child to, number of siblings, math scores, English grades, science grades, Indonesian language scores, test scores, and majors recommendations. After testing using K-Fold Cross Validation and Confusion Matrix for evaluation and validation of results with Euclidean Distance calculation, the best value of k is k=3 which produces accuracy: 97.11%, precision: 96.82%, recall: 98.33%, and AUC: 0.951.

Keywords: Majoring new students, K-Nearest Neighbor (K-NN) Algorithm, K-Fold Cross Validation, Confusion Matrix

MOTTO

“To Infinity and Beyond”

“Menuju tak terbatas dan melampauinya”

(Buzz Lightyear)

“Sesungguhnya Allah tidak akan mengubah keadaan suatu kaum sehingga mereka merubah keadaan yang ada pada diri mereka sendiri”

(Q. S Ar-Ra’d: 11)

“Dan mintalah pertolongan dengan sabar dan sholat”

(Q. S Al-Baqarah: 45)

PERSEMBAHAN

Skripsi ini saya persembahkan kepada :

1. Kedua orang tuaku Bapak Konjim dan Ibu Rosyidah yang telah merawat, menyayangi, dan selalu mendo'akanku.
2. Kakakku Haris dan Nunung yang sudah mendo'akan, membantu dan memberikan semangat serta adikku Amsa.
3. Teman-teman yang selalu membantu dan memberikan semangat yaitu Tuvie, Anang, Virgiawan, Ilyas, dan Yusuf.
4. Teman-teman TIF Angkatan 2018.
5. Teman-teman ORSHA UNISNU
6. Kepada jiwa raga, otak dan fikiranku.

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan atas kehadiran Allah SWT yang telah memberikan rahmat, taufiq, dan hidayah-Nya, sehingga peneliti dapat menyelesaikan penelitian dengan judul: “Optimasi Parametr K Pada Algoritma K-NN Untuk Klasifikasi Jurusan Peserta Didik Baru Di MA Darul Hikmah Menganti” dengan baik.

Peneliti menyadari dalam menyelesaikan Skripsi ini, terdapat keterbatasan dari pengetahuan yang dimiliki. Oleh karena itu, tanpa keterlibatan dan sumbangsih dari berbagai pihak, sulit bagi peneliti untuk menyelesaikan penyusunan Skripsi ini. Maka, dengan segala kerendahan hati serta rasa syukur yang setinggi-tingginya kepada Allah SWT dan Nabi Muhammad SAW atas izin dan karunia-Nya penulis dapat menyelesaikan skripsi ini. Peneliti juga mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Rektor Universitas Islam Nahdlatul Ulama (UNISNU) Jepara, Bpk. Dr. H. Sa’dullah Assaidi, M.Ag.,
2. Dekan Fakultas Sains dan Teknologi Universitas Islam Nahdlatul Ulama (UNISNU) Jepara, Bpk. Dias Prihatmoko ST., M.Eng.
3. Ketua Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Nahdlatul Ulama (UNISNU) Jepara, Bpk. Gentur Wahyu Nyipto Wibowo, S.Kom, M.Kom.,
4. Pembimbing I, Bpk. Harminto Mulyo, S.Kom., M.Kom. dan pembimbing II, Nur Aeni Widiastuti, S.Pd., M.Kom. yang telah membimbing peneliti dengan penuh kesabaran dan tanggungjawab dalam penyusunan skripsi hingga selesai.
5. Bapak dan Ibu Dosen Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Nahdlatul Ulama (UNISNU) Jepara.
6. Pihak MA Darul Hikmah Menganti yang telah memberikan izin kepada peneliti untuk melaksanakan penelitian hingga selesai.

7. Kedua orang tuaku Bapak Konzim dan Ibu Rosidah yang telah merawat, menyayangi, dan selalu mendo'akanku.
8. Kakakku Haris dan Nunung yang sudah mendo'akan, membantu dan memberikan semangat serta adikku Amsa.
9. Teman-teman NGEKLEK Tuvie, Anang, Virgiawan, Ilyas, dan Yusuf yang telah membantu dan memberikan semangat.
10. Teman Program Studi Teknik Informatika angkatan tahun 2018 Fakultas Sains dan Teknologi Universitas Islam Nahdlatul Ulama (UNISNU) Jepara.
11. Teman-teman UKM ORSHA khususnya Angkatan 2018
12. Semua pihak baik secara langsung maupun tidak langsung yang telah memberikan bantuan fisik maupun psikis, dalam menyelesaikan skripsi ini yang tidak dapat disebutkan satu per satu.

Peneliti juga menyadari bahwa dalam penulisan Skripsi ini masih jauh dari kata sempurna. Maka, kritik dan saran yang bersifat membangun sangat dibutuhkan guna menyempurnakan Skripsi ini nantinya. Peneliti berharap Skripsi ini dapat memberikan manfaat bagi peneliti sendiri maupun pembaca yang melihat dan membacanya.

Jepara, 5 Agustus 2022

Peneliti,



Maulana Ashar

NIM. 181240000759

DAFTAR ISI

HALAMAN JUDUL	i
PERSETUJUAN PEMBIMBING	ii
PENGESAHAN	iii
PERNYATAAN KEASLIAN	iv
ABSTRAK	iv
MOTTO	vi
PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Batasan Masalah	6
1.3 Rumusan Masalah	6
1.4 Tujuan	7
1.5 Manfaat Penelitian	7
1.6 Sistematika Penulisan	7
BAB II LANDASAN TEORI	10
2.1 Tinjauan Studi	10
2.2 Tinjauan Pustaka	14
2.2.1 Penjurusan Siswa Baru	14
2.2.2 Data Mining	14
2.2.3 Algoritma K-Nearest Neighbor (KNN)	19
2.2.4 Cross validation	26
2.2.5 Confusion Matrix	27
2.2.6 Kurva Receiver Operating Characteristics (ROC)	29
2.2.7 Rapidminer	30
2.3 Kerangka pemikiran	32

BAB III METODE PENELITIAN	33
3.1 Desain Penelitian	33
3.2 Teknik Pengumpulan Data	35
3.3 Lokasi Penelitian.....	36
3.4 Pengelolaan Data Awal.....	36
3.5 Metode Yang Diusulkan	38
3.6 Pengujian Model	38
3.7 Evaluasi dan Validasi Hasil.....	39
BAB IV HASIL DAN PEMBAHASAN	40
4.1 Hasil Penelitian	40
4.2 Pembahasan	42
4.2.1 Pengumpulan Data.....	42
4.2.2 Pengelolaan Data Awal.....	42
4.2.3 Pengujian Model K-Fold Cross Validation	51
4.2.4 Algoritma K-Nearest Neighbor (K-NN).....	58
4.2.5 Evaluasi dan Validasi.....	68
4.2.6 Kurva ROC (Receiver Operating Characteristics).....	71
BAB V PENUTUP	74
5.1 Kesimpulan	74
5.2 Saran	74
DAFTAR PUSTAKA	75
LAMPIRAN-LAMPIRAN	79

DAFTAR TABEL

Tabel 2. 1 Contoh Dataset.....	22
Tabel 2. 2 Hasil Euclidean Distance	25
Tabel 2. 3 Urutan jarak dari yang terkecil.....	25
Tabel 2. 4 Urutan Euclidean Distance Dataset.....	25
Tabel 2. 5 Penentuan Kategori Data Testing	26
Tabel 2. 6 Contoh pembagian dataset	27
Tabel 2. 7 Contoh pembagian dataset	27
Tabel 2. 8 Hasil pengujian akurasi.....	27
Tabel 2. 9 Confusion Matrix	28
Tabel 2. 10 Contoh perhitungan Confusion Matrix	29
Tabel 2. 11 klasifikasi AUC.....	30
 Tabel 3. 1 Kriteria Variabel	 37
 Tabel 4. 1 K-fold cross validation.....	 41
Tabel 4. 2 Confusion Matrix	41
Tabel 4. 3 Tipe Data.....	47
Tabel 4. 4 Hasil Akurasi 10-fold cross validation.....	57
Tabel 4. 5 Dataset Penelitian.....	58
Tabel 4. 6 Evaluasi perhitungan jarak.....	60
Tabel 4. 7 Hasil Perhitungan Euclidean Distance	66
Tabel 4. 8 Penentuan kelas mayoritas	67
Tabel 4. 9 Hasil Klasifikasi Data Testing dengan Microsoft Excel.....	68
Tabel 4. 10 Performance Vector	69

DAFTAR GAMBAR

Gambar 1. 1 Jumlah siswa-siswi kelas X pertahun	2
Gambar 1. 2 Alur penjurusan siswa baru	3
Gambar 2. 1 Proses Knowledge Discovery In Database (KDD) [12].....	15
Gambar 2. 2 Flowchart K-Nearest Neighbor	20
Gambar 2. 3 Tampilan RapidMiner	32
Gambar 2. 4 Kerangka Pemikiran.....	33
Gambar 3. 1 Flowchart Penelitian.....	33
Gambar 4. 1 Data Awal Sebelum Diseleksi Atribut	43
Gambar 4. 2 Data Awal Setelah Diseleksi Atribut	44
Gambar 4. 3 Permodelan Seleksi Atribut.....	45
Gambar 4. 4 Parameters Select Attributes	45
Gambar 4. 5 Seleksi Atribut.....	46
Gambar 4. 6 Hasil Seleksi Atribut	46
Gambar 4. 7 Hasil Penyusunan dan Pengkategorian Data	47
Gambar 4. 8 Mengubah Tipe Data Pada Format Kolom	48
Gambar 4. 9 Permodelan Replace Missing Values	49
Gambar 4. 10 Pemilihan Parameters Replace Missing Values	49
Gambar 4. 11 Hasil ExampleSet Pembersihan Data	50
Gambar 4. 12 Hasil Statistic replace missing values atribut	50
Gambar 4. 13 Permodelan K-Fold Cross Validation	51
Gambar 4. 14 Permodelan 10-fold cross validation.....	52
Gambar 4. 15 Permodelan Cross Validation.....	52
Gambar 4. 16 Split data.....	53
Gambar 4. 17 Parameters Algoritma K-NN.....	53
Gambar 4. 18 Kolom Training	54

Gambar 4. 20 Hasil pengujian k=1	54
Gambar 4. 21 Hasil pengujian k=3	54
Gambar 4. 22 Hasil pengujian k=5	55
Gambar 4. 23 Hasil pengujian k=7	55
Gambar 4. 24 Hasil pengujian k=9	55
Gambar 4. 25 Hasil pengujian k=11	56
Gambar 4. 26 Hasil pengujian k=13	56
Gambar 4. 27 Hasil pengujian k=15	56
Gambar 4. 28 Hasil pengujian k=17	57
Gambar 4. 29 Hasil pengujian k=19	57
Gambar 4. 30 Hasil Kelas Mayoritas RapidMiner.....	67
Gambar 4. 31 Performance accuracy	70
Gambar 4. 32 Performance precision.....	70
Gambar 4. 33 Performance Recall	71
Gambar 4. 34 Kurva ROC (Receiver Operating Characteristic).....	72

DAFTAR LAMPIRAN

Lampiran 1 Buku Bimbingan Skripsi	80
Lampiran 2 Surat ijin penelitian.....	82
Lampiran 3 Formulir dan Hasil Wawancara.....	83
Lampiran 4 Foto – Foto	85
Lampiran 5 Data Training dan Data Testing.....	86