

Perbandingan Metode *K-Nearest Neighbour* dan *Fuzzy K-Nearest Neighbour* dalam Klasifikasi Data

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika*



Oleh:

**Rio Andika Saputra
09021281419048**

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2020**



LEMBAR PENGESAHAN TUGAS AKHIR

Perbandingan Metode *K-Nearest Neighbour* dan *Fuzzy K-Nearest Neighbour*
Dalam Klasifikasi Data

Oleh :

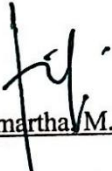
Rio Andika Saputra

NIM : 09021281419048

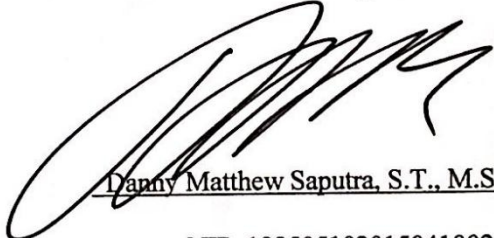
Palembang, 3 Januari 2020

Pembimbing I

Pembimbing II,


Rifkie Primartha, M.T.

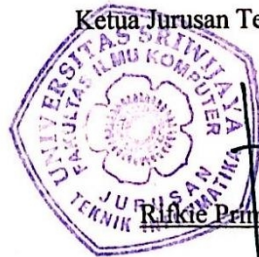
NIP. 197706012009121004


Danny Matthew Saputra, S.T., M.Sc.

NIP. 198505102015041002

Mengetahui,

Ketua Jurusan Teknik Informatika




Rifkie Primartha, M.T.

NIP. 197706012009121004



TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada Hari tanggal 31 Desember 2019 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Rio Andika Saputra

NIM : 09021281419048

Judul Skripsi : Perbandingan Metode *K-Nearest Neighbour* dan *Fuzzy K-Nearest Neighbour* Dalam Klasifikasi Data.

1. Pembimbing I,

Rifkie Primartha, M.T.

NIP. 197706012009121004

2. Pembimbing II,

Danny Matthew Saputra, S.T., M.Sc.

NIP. 198505102015041002

3. Penguji I,

Rusdi Efendi, M.Kom.

NIP. 198201022015109191

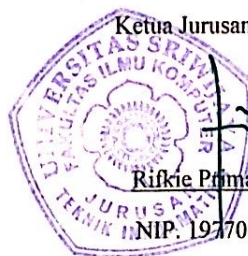
4. Penguji II,

Desty Rodiah, M.T.

NIP.

Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.

NIP. 197706012009121004



HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Rio Andika Saputra

NIM : 09021281419048

Program Studi : Teknik Informatika

Judul Skripsi : Perbandingan Metode *K-Nearest Neighbour* dan *Fuzzy K-Nearest Neighbour* Dalam Klasifikasi Data

Hasil Pengecekan Software *iThenticate/Turnitin* : 15%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 3 Januari 2020

METERAI
TEMPEL
6E29BAHF7953B797B
6000
ENAM RIBURUPIAH

Rio Andika Saputra

NIM. 09021281419048



MOTTO DAN PERSEMBAHAN

Motto :

While you're young and you got a chance to do it, DO IT.

Kupersembahkan karya tulis ini

kepada :

- Orang tua
- Keluarga besar
- Fakultas Ilmu Komputer
Universitas Sriwijaya



COMPARISON OF K-NEAREST NEIGHBOUR AND FUZZY K-NEAREST NEIGHBOUR METHOD IN DATA CLASSIFICATION

Oleh :

Rio Andika Saputra

09021281419048

ABSTRACT

K-NN is such an effective and the most commonly used method for data classification. This method also believed as one of the most popular and easiest method to be implemented in this subject. But, there is also a method that has a lot of similarities with this one (K-NN) and known for its ability to determine an object category also by its weight, this method called Fuzzy K-NN. Fuzzy K-Nearest Neighbour basically is just the K-NN with Fuzzy theory in it. The only difference between this two is the ability of Fuzzy K-NN to determines the membership value in each individuals in data classification process. in this study, writer tried to examine the affects of weights (leads to membership value) on the accuracy of the data classification (in each K) by using iris and blood transfusion data as the objects. Based on the research results, it was found that the best accuracy for both K-NN and Fuzzy K-NN are 98,04% for iris data, with $K = 1$ in K-NN and $K = 13$ in Fuzzy K-NN. Meanwhile, in blood transfusion data classification the best results for both methods shows that K-NN has a better accuracy (80,54%) than the Fuzzy K-NN (78,95%) in $K = 20$.

Keywords : Classification, K-nearest neighbour, Fuzzy K-nearest neighbour, weights, Membership value.

Inderalaya, 3 january 2020

Pembimbing I

Pembimbing II,

Rifkie Primartha, M.T.

NIP. 197706012009121004

Danny Matthew Saputra, S.T., M.Sc.

NIP. 198505102015041002



KATA PENGANTAR

Assalamu'alaikum Wr.Wb

Puji syukur kepada Allah SWT karena atas izin-Nya penulis dapat menyelesaikan Laporan Skripsi dengan judul "Perbandingan Metode *K-Nearest Neighbour* dan *Fuzzy K-Nearest Neighbour* dalam Klasifikasi Data" sebagai salah satu syarat menyelesaikan program S-1 pada Fakultas Ilmu Komputer, Program Studi Teknik Informatika Universitas Sriwijaya.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah memberikan dukungan, bimbingan, dan motivasi kepada penulis untuk menyelesaikan tugas akhir ini, yaitu kepada :

1. Allah SWT yang telah memberikan nikmat kemudahan serta kelancaran sehingga penulis dapat menyelesaikan laporan skripsi ini,
2. Kedua orang tua dan keluarga tercinta yang telah memberikan dukungan moril, materil, dan doa yang tiada hentinya,
3. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya,
4. Bapak Rifkie Primartha, S.T., M.T. Selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya, serta Pembimbing I Tugas Akhir yang telah membimbing, dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir,
5. Bapak Danny Matthew Saputra, S.T., M.Sc. Selaku Pembimbing II Tugas Akhir yang telah membimbing, dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir,
6. Bapak Osvari Arsalan, S.Kom., M.T. Selaku Pembimbing Akademik yang telah membimbing, mengarahkan, dan memberikan motivasi penulis dalam proses perkuliahan,



7. Bapak Rusdi Efendi, M.Kom. Selaku Penguji I yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir,
8. Ibu Desty Rodiah, M.T. Selaku Penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir,
9. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya,
10. Kak Ricky dan Kak Hafez serta seluruh staff tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan,
11. Seluruh teman-teman jurusan Teknik Informatika dan teman-teman yang telah membantu dan mendukung penulis selama menempuh pendidikan di Fakultas Ilmu Komputer Universitas Sriwijaya,
12. Dan yang spesial pake telur Asya Maharani yang telah memberikan semangat serta menemani penulis selama proses pengerjaan Tugas Akhir.

Penulis menyadari bahwa didalam pembuatan laporan ini terdapat banyak kekurangan, oleh karena itu penulis mengharapkan saran, kritik dan koreksi yang sifatnya membangun dan bisa digunakan sebagai masukan dikemudian hari. Akhir kata penulis berharap semoga laporan tugas akhir ini dapat bermanfaat bagi kita semua. Terima kasih.

Wassalamu'alaikum Wr. Wb

Palembang, 3 Januari 2020



Rio Andika Saputra



DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERSETUJUAN KOMISI PENGUJI	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iv
HALAMAN MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABLE	xiv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN	xvii
BAB 1 PENDAHULUAN	
1.1 Pendahuluan.....	I-1
1.2 Latar Belakang Masalah.....	I-1
1.3 Rumusan Masalah.....	I-3
1.4 Tujuan Penelitian.....	I-3
1.5 Manfaat Penelitian.....	I-3
1.6 Batasan Masalah.....	I-3
1.7 Sistematika Tulisan.....	I-5
BAB II KAJIAN LITERATUR	
2.1 Pendahuluan.....	II-6
2.2 Klarifikasi.....	II-6
2.3 Data.....	II-7
2.3.1 Normalisasi Data.....	II-8
2.4 K-Nearest Neighbor.....	II-8
2.5 Fuzzy K-Nearest Neighbor.....	II-9



2.6 K-Fold Cross Validation	II – 10
2.7 Confusion Matrix.....	II – 12
2.8 Penelitian Lain Yang Relevan.....	II – 13
2.8.1 (Satria Dwi Nugraha, Rekyan Regasari Mardi Putri, Randy Cahya Wihandika, 2017) : Penerapan <i>Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentuksn Status Gizi Balita</i> , Fakultas Ilmu Komputer, Universitas Brawijaya, Indonesia.....	I.I-13
2.8.2 (Andhika Satria Pria Anugerah, Indriati, Candra Dewi, 2018): <i>Implementasi Algoritme Fuzzy K-Nearest Neighbor untuk Penentuan Lulus Tepat Waktu</i> , Fakultas Ilmu Komputer, Universitas Brawijaya, Indonesia.....	II-14

BAB III METODOLOGI PENELITIAN

3.1 Pendahuluan.....	III – 15
3.2 Data.....	III – 15
3.2.1 Jenis dan Sumber Data.....	III – 15
3.2.2 Metode Pengumpulan Data.....	III – 15
3.3 Tahapan Penelitian.....	III – 16
3.3.1 Menetapkan Kerangka Kerja/Framework.....	III – 16
3.3.2 Menetapkan Kriteria Pengujian.....	III – 18
3.3.3 Menetapkan Format Data Pengujian.....	III – 18
3.3.4 Menentukan alat yang digunakan Dalam Pelaksanaan Penelitian.....	III – 19
3.3.5 Melakukan Pengujian Penelitian.....	III – 19
3.3.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan.....	II I- 20
3.4 Metode Pengembangan Perangkat Lunak.....	III – 20
3.4.1 Fase Insepsi.....	III – 21
3.4.2 Fase Elaborasi.....	III – 21
3.4.3 Fase Konstruksi.....	III – 21
3.4.4 Fase Transisi.....	III – 22
3.5 Penjadwalan Perencanaan.....	III – 22



BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1 Pendahuluan.....	IV-23
4.2 Fase Insepsi.....	IV-23
4.2.1 Analisis Perangkat Lunak.....	IV-23
4.2.1.1 Deskripsi Umum Perangkat Lunak.....	IV-23
4.2.1.2 Analisis Kebutuhan Perangkat Lunak.....	IV-23
4.2.1.3 Analisis Prapengolahan Data.....	IV-24
4.2.1.4 Normalisasi Data.....	IV-27
4.2.1.5 Analisis Validasi dengan K-fold Cross validation	IV-30
4.2.1.6 Analisis Klasifikasi dengan KNN.....	IV-31
4.2.1.7 Analisis Klasifikasi dengan FKNN.....	IV-33
4.2.2 Desain Perangkat Lunak.....	IV-34
4.2.2.1 Use Case Diagram	IV-34
4.2.2.2 Activity Diagram.....	IV-40
4.3 Fase Elaborasi.....	IV-44
4.3.1 Arsitektur Perangkat Lunak.....	IV-44
4.3.2 Perancangan Data.....	IV-44
4.3.3 Perancangan Antar Muka.....	IV-44
4.3.4 Perancangan <i>Entity Relationship Diagram</i> (ERD).....	IV-46
4.3.5 Perancangan Diagram <i>Sequence</i>	IV-46
4.3.6 Perancangan Diagram Kelas.....	IV-49
4.4 Fase Konstruksi.....	IV-52
4.4.1 Implementasi Kelas.....	IV-52
4.4.2 Implementasi Antarmuka.....	IV-55
4.5 Fase Transisi.....	IV-56



4.5.1 Lingkungan Pengujian.....	IV-56
4.5.2 Rencana Pengujian.....	IV-57
4.5.2.1 Rencana Pengujian Use Case Validasi Menggunakan Cross Validation.....	IV-57
4.5.2.2 Pengujian <i>Use Case</i> Muat Data.....	IV-57
4.5.2.3 Rencana Pengujian <i>Use Case</i> Klasifikasi Menggunakan KNN.....	IV-58
4.5.2.4 Rencana Pengujian <i>Use Case</i> Klasifikasi Menggunakan FKNN.....	IV-58
4.5.3 Kasus Uji.....	IV-58
4.5.3.1 Pengujian <i>Use Case</i> Validasi Menggunakan Cross Validation.....	IV-59
4.5.3.2 Pengujian <i>Use Case</i> Muat Data.....	IV-59
4.5.3.3 Pengujian <i>Use Case</i> Klasifikasi Menggunakan KNN....	IV-60
4.5.3.4 Pengujian <i>Use Case</i> Klasifikasi Menggunakan FKNN..	IV-61
4.6 Kesimpulan.....	IV-63
BAB V HASIL DAN ANALISIS PENELITIAN	
5.1 Pendahuluan.....	V-64
5.2 Hasil Percobaan Penelitian.....	V-64
5.2.1 Percobaan Menggunakan Data <i>Iris</i>	V-64
5.2.1.1 Percobaan K = 1.....	V-64
5.2.1.2 Percobaan K = 4.....	V-66
5.2.1.3 Percobaan K = 7	V-67
5.2.1.4 Percobaan K = 9	V-68
5.2.1.5 Percobaan K = 13	V-69
5.2.1.6 Percobaan K = 15	V-70
5.2.1.7 Percobaan K = 20.....	V-71



5.2.1.8 Percobaan $K = 30$	V-72
5.2.1.9 Hasil analisa performa klasifikasi pada data <i>iris</i>	V-73
5.2.2 Percobaan Menggunakan Data <i>Blood Transfusion</i>	V-73
5.2.2.1 Percobaan $K = 1$	V-73
5.2.2.2 Percobaan $K = 4$	V-74
5.2.2.3 Percobaan $K = 7$	V-75
5.2.2.4 Percobaan $K = 9$	V-76
5.2.2.5 Percobaan $K = 13$	V-77
5.2.2.6 Percobaan $K = 15$	V-78
5.2.2.7 Percobaan $K = 20$	V-79
5.2.2.8 Percobaan $K=30$	V-80
5.2.2.9 Hasil analisa performa klasifikasi pada data <i>iris</i>	V-81
5.3 Analisis Hasil Penelitian.....	V-82
5.4 Kesimpulan	V-83
BAB VI KESIMPULAN DAN SARAN	
6.1 Pendahuluan.....	VI-84
6.2 Kesimpulan.....	VI-84
6.3 Saran.....	VI-85

DAFTAR PUSTAKA



DAFTAR TABEL

Tabel III-1. Rancangan Tabel Format Data Pengujian.....	III-19
Tabel III-2 Rancangan Tabel Hasil Analisa Performa Algoritma Klasifikasi..	III-20
Tabel III-3 Penjadwalan Penelitian.....	III-22
Tabel IV – 1 Kebutuhan Fungsional.....	IV-25
Tabel IV – 2 Kebutuhan Non Fungsional.....	IV-25
Tabel IV-3. Data <i>Iris</i> Sebelum Prapengolahan.....	IV-27
Tabel IV-4. Data <i>Iris</i> Setelah Normalisasi.....	IV-28
Tabel IV-5 Representasi K-fold = 2 pada data uji pertama.....	IV-29
Tabel IV-6 Representasi K-fold = 2 pada data uji kedua.....	IV-29
Tabel IV-7 Representasi simulasi data uji pada KNN.....	IV-30
Tabel IV-8 Representasi simulasi data latih pada KNN.....	IV-30
Tabel IV-8 Simulasi jarak data uji terhadap data latih.....	IV-31
Tabel IV-9 Simulasi hitung jarak terhadap bobot dan nilai keanggotaan.....	IV-32
Tabel IV – 10. Definisi Aktor <i>Use Case</i>	IV-33
Tabel IV – 11. Definisi <i>Use Case</i>	IV-34
Tabel IV-12. Skenario <i>Use Case</i> Menyimpan Data.....	IV-35
Tabel IV-13. Skenario <i>Use Case</i> Validasi dengan Cross Validation.....	IV-36
Tabel IV – 14. Skenario <i>Use Case</i> Menggunakan Algoritma KNN.....	IV-37
Tabel IV – 15. Skenario <i>Use Case</i> Menggunakan Algoritma FKNN.....	IV-38
Tabel IV –16 . Implementasi Kelas.....	IV-51
Tabel IV –17 . Rencana Pengujian <i>Use Case</i> Validasi Menggunakan Cross Validation.....	IV-56
Tabel IV–18 Rencana Pengujian <i>Use Case</i> Muat Data.....	IV-56



Tabel IV-19 Rencana Pengujian <i>Use Case</i> Klasifikasi Menggunakan KNN	IV-57
Tabel IV-20 Rencana Pengujian <i>Use Case</i> Klasifikasi Menggunakan FKNN.....	IV-57
Tabel IV-21 Pengujian <i>Use Case</i> Validasi Menggunakan Cross Validation.....	IV-58
Tabel IV-22 Rencana Pengujian <i>Use Case</i> Muat Data.....	IV-58
Tabel IV-23 Rencana Pengujian <i>Use Case</i> Klasifikasi Menggunakan KNN.....	IV-58
Tabel IV-24 Rencana Pengujian <i>Use Case</i> Klasifikasi Menggunakan FKNN.....	IV-59
Tabel V-1. Pengujian Klasifikasi K = 1 (%).....	V-61
Tabel V-2. Pengujian Klasifikasi K = 4 (%).....	V-63
Tabel V-3 Pengujian Klasifikasi K = 7 (%).....	V-64
Tabel V-4. Pengujian Klasifikasi K = 9 (%).....	V-65
Tabel V-5. Pengujian Klasifikasi K = 13 (%).....	V-66
Tabel V-6. Pengujian Klasifikasi K = 15 (%).....	V-66
Tabel V-7. Pengujian Klasifikasi K = 20 (%).....	V-67
Tabel V-8. Pengujian Klasifikasi K = 30 (%).....	V-68
Tabel V-9. Performa KNN dan FKNN pada <i>iris</i>	V-69
Tabel V-10. Pengujian akurasi K = 1 (%).....	V-70
Tabel V-11. Pengujian akurasi K = 4 (%).....	V-71
Tabel V-12. Pengujian akurasi K = 7 (%).....	V-72
Tabel V-13. Pengujian akurasi K = 9 (%).....	V-73
Tabel V-14. Pengujian akurasi K = 13 (%).....	V-74
Tabel V-15. Pengujian akurasi K = 15 (%).....	V-75
Tabel V-16. Pengujian akurasi K = 20 (%)	V-76
Tabel V-17. Pengujian akurasi K = 30 (%).....	V-77
Tabel V-18. Performa KNN dan FKNN pada <i>blood transfusion</i>	V-78



DAFTAR GAMBAR

Gambar II-1 Proses Klasifikasi.....	II-6
Gambar II-2. Pembagian dataset <i>K-fold Cross Validation</i>	II-9
Gambar II-3 Ilustrasi penentuan golongan data dengan K di KNN.....	II-9
Gambar II-3 Confusion Matix.....	II-12
Gambar III-1 Kerangka Kerja.....	III-16
Gambar IV – 1. Diagram <i>Use Case</i>	IV-33
Gambar IV-2. Muat Data.....	IV-39
Gambar IV-3. Diagram aktivitas Klasifikasi dengan <i>cross validation</i>	IV-40
Gambar IV-4. Menggunakan Algoritma KNN.....	IV-41
Gambar IV-5. Diagram aktivitas menggunakan algoritma FKNN.....	IV-42
Gambar IV-6. Panel tampilan pertama antarmuka perangkat lunak.....	IV-44
Gambar IV-7 . Panel untuk memasukkan data ke dalam perangkat lunak.....	IV-44
Gambar IV-8 . Panel klasifikasi algoritma.....	IV-45
Gambar IV-9 . Validasi menggunakan <i>cross validation</i>	IV-46
Gambar IV-10. Muat Data.....	IV-46
Gambar IV-11 . Menggunakan Algoritma KNN.....	IV-47



BAB I PENDAHULUAN

1.1 Pendahuluan

Latar belakang masalah, rumusan masalah, tujuan, manfaat penelitian serta batasan masalah dibahas dalam Bab ini. Penjelasan akan diawali dengan pembahasan singkat mengenai pengertian data mining yang disertai juga dengan pembahasan dari metode KNN dan FKNN yang digunakan pada klasifikasi data untuk dilakukan perbandingan pada kedua metode yang digunakan.

1.2 Latar Belakang

Tanpa disadari setiap individu yang hidup di era modern seperti sekarang ini akan terus menghasilkan serta berurusan dengan data dalam menjalani kegiatan sehari-hari. Pasalnya, manusia telah hidup di era data yang dimana penggunaan data digital telah menjangah hampir seluruh aspek kehidupan manusia mulai dari aspek olahraga, kesehatan, pendidikan, perkebunan bahkan aspek krusial seperti pemerintahan. Sedikitnya, manusia menghasilkan sekitar 2.5 triliun byte data di setiap harinya. Kumpulan data yang masif pada setiap aspek kehidupan ini, dimanfaatkan oleh sebagian besar perusahaan dan individu tertentu untuk menemukan korelasi dan ilmu pengetahuan yang berguna didalamnya, karena pada dasarnya data juga memberikan pengaruh pada valuasi dan reputasi suatu perusahaan. Keputusan bagaimana perusahaan menggunakan data bisa memberikan dampak pada kesuksesan perusahaan modern itu sendiri. Start up Gojek misalnya, melakukan analisis terhadap data pelanggannya untuk kemudian dimanfaatkan sebagai *tools* untuk memperkirakan selera pelanggan, rekomendasi makanan dan mengatur preferensi pengemudi dalam memilih rute yang dilalui. Salah satu teknik olah data yang dapat digunakan untuk menganalisis data adalah data mining.

Data mining merupakan suatu proses yang menggunakan teknik yang cerdas dalam mengekstraksi pola-pola pengetahuan yang memiliki potensi kegunaan didalamnya (osmar,1999). Beberapa teknik data mining yang sering digunakan



antara lain yaitu klasifikasi, prediksi, estimasi, asosiasi dan pengelompokan. Dari beberapa macam teknik tersebut, teknik yang paling sering digunakan adalah teknik klasifikasi (Kaur dan Grewal, 2016).

Klasifikasi adalah salah satu teknik pada penerapan *data mining*, yang dimana teknik ini digunakan untuk menempatkan suatu objek ke dalam suatu kategori/kelas yang sudah didefinisikan sebelumnya (Emerensye, 2013). Klasifikasi juga merupakan salah satu topik utama dalam data mining atau machine learning. Teknik ini bekerja pada data yang memiliki kelas label atau target sehingga algoritma-algoritma untuk menyelesaikan masalah klasifikasi dikategorisasikan ke dalam supervised learning atau pembelajaran yang diawasi. KNN adalah algoritma dari machine learning yang pada umumnya digunakan pada penklasifikasian data. KNN dianggap sebagai algoritma klasifikasi yang sangat efisien, yang dimana algoritma ini mudah untuk diimplementasikan, mudah untuk digunakan dan juga efektif dalam menghadapi data yang noise (Rani dan Vashishtha 2017). Namun, algoritma KNN ini kurang efektif untuk klasifikasi, karena pada dasarnya KNN melakukan pengklasifikasian data berdasarkan data mayoritas kelas yang ada pada sejumlah K tetangga terdekat dari data baru tersebut (Anugerah, Indriati dan Dewi, 2018). Oleh karena itu, Teori Fuzzy dikombinasikan dengan algoritma KNN (FKNN) untuk klasifikasi data.

FKNN menggunakan teori himpunan fuzzy untuk menggeneralisir teori knn klasik dengan mendefinisikan nilai keanggotaan sebuah data pada masing-masing kelas K terdekat. Penerapan Fuzzy KNN dalam menentukan status gizi balita dengan menggunakan 4 faktor internal yaitu jenis kelamin, umur, berat badan dan tinggi badan yang menggunakan 160 data latih, 32 data uji dan nilai $K = 4$ menghasilkan akurasi FKNN sebesar 82% dan untuk KNN sebesar 71% menggunakan K-fold cross-validation (Nugraha, Putri dan Wihandika, 2017). Implementasi algoritma FKNN dan KNN untuk penentuan lulus tepat waktu dengan menggunakan parameter data kelulusan berupa indeks prestasi semester 1 hingga indeks prestasi semester 14 dimana data tersebut memiliki 2 label yaitu tepat waktu dan tidak tepat waktu berjumlah 189 data menghasilkan akurasi terbaik KNN sebesar 96% dan akurasi terbaik FKNN sebesar 98% (Anugerah, Indriati dan Dewi, 2018).



Berdasarkan uraian di atas, maka pada penelitian ini peneliti akan membandingkan algoritma KNN dan FKNN pada 2 objek data yang berbeda dalam melakukan klasifikasi.

1.3 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah :

1. Bagaimana pengaruh dari teori himpunan fuzzy terhadap akurasi algoritma KNN untuk klasifikasi data?
2. Bagaimana metode KNN dan FKNN melakukan klasifikasi terhadap data?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini sebagai berikut :

1. Mengukur metode mana yang memiliki tingkat keakuratan yang paling tinggi dalam klasifikasi data.
2. Mengukur akurasi di setiap K pada masing-masing metode dalam klasifikasi data.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini sebagai berikut :

1. Mengetahui performa teori himpunan fuzzy terhadap algoritma KNN dalam melakukan klasifikasi data.
2. Sebagai referensi penelitian yang digunakan selanjutnya.

1.6 Batasan Masalah

Batasan masalah yang akan dibahas di penelitian ini sebagai berikut :

1. Data yang digunakan merupakan data berjenis sekunder yaitu data iris dan *blood transfusion* yang diperoleh dari UCI Repository (<http://archive.ics.uci.edu/ml/>).
2. Metode validasi yang digunakan adalah *cross-validation*.
3. *K-fold* yang digunakan berjumlah 10 fold.
4. Percobaan menggunakan 8 buah jenis K tetangga terdekat pada algoritma KNN dan fuzzy KNN.
5. Bobot pangkat dari metode fuzzy KNN bernilai 2.



1.7 Sistematika Tulisan

Sistematika penulisan skripsi adalah uraian singkat bab perbab.

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan, masalah/ruang lingkup, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian, seperti definisi dan penjelasan konsep seperti klasifikasi, data yang digunakan, Normalisasi data, *K-fold cross validation*, *Equalization*, *K-Nearest Neighbor*, *Fuzzy K-Nearest Neighbor*, *Confussion matrix* serta dilengkapi dengan kesimpulan dari penelitian yang relevan.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dibahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan dibahas mengenai perancangan dan implementasi program perbandingan *K-Nearest Neighbor* dan *Fuzzy K-Nearest Neighbor* dalam klasifikasi data..

BAB V. HASIL DAN ANALISIS PENELITIAN

Hasil dari pengujian didasarkan atas langkah-langkah yang telah direncanakan sebelumnya. Analisis disajikan sebagai dasar dari kesimpulan yang diambil.



BAB VI. KESIMPULAN DAN SARAN

Kesimpulan dari semua uraian pada bab-bab sebelum ini beserta saran dalam penerapan KNN dan FKNN dalam melakukan klasifikasi terhadap data.

DAFTAR PUSTAKA

- [1] O. R. Za, S. Pole, and C. Science, "Chapter I : Introduction to Data Mining," pp. 1–15, 1999.
- [2] S. Kaur and A. K. Grewal, "A REVIEW PAPER ON DATA MINING CLASSIFICATION TECHNIQUES FOR," pp. 1334–1338, 2016.
- [3] P. Rani and J. Vashishtha, "An Appraise of KNN to the Perfection," vol. 170, no. 2, pp. 13–17, 2017.
- [4] P. Anugerah, Indriati, and C. Dewi, "Implementasi Algoritme Fuzzy K-Nearest Neighbor untuk Penentuan Lulus Tepat Waktu (Studi Kasus : Fakultas Ilmu Komputer Universitas Brawijaya)," vol. 2, no. 4, pp. 1726–1732, 2018.
- [5] S. D. Nugraha, R. R. M. Putri, and R. C. Wihandika, "Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita," no. February, 2018.
- [6] D. Nofriansyah, "Konsep Data Mining VS Sistem Pendukung Keputusan," 2015.
- [7] P. Rani and J. Vashishtha, "An Appraise of KNN to the Perfection," vol. 170, no. 2, pp. 13–17, 2017.
- [8] P. Anugerah, Indriati, and C. Dewi, "Implementasi Algoritme Fuzzy K-Nearest Neighbor untuk Penentuan Lulus Tepat Waktu (Studi Kasus : Fakultas Ilmu Komputer Universitas Brawijaya)," vol. 2, no. 4, pp. 1726–1732, 2018.
- [9] S. D. Nugraha, R. R. M. Putri, and R. C. Wihandika, "Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita," no. February, 2018.