

## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan segala Rahmat, Karunia dan Hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul : **“Perbandingan K-Nearest Neighbor dan Fuzzy K-Nearest Neighbor pada Diagnosis Penyakit Diabetes Melitus”**.

Skripsi ini diajukan sebagai syarat ujian seminar skripsi dalam rangka untuk memperoleh gelar Sarjana Komputer di Program Studi Teknik Informatika dan Ilmu Komputer Program Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya Malang. Atas terselesaikannya skripsi ini, penulis mengucapkan terima kasih kepada :

1. Drs. Achmad Ridok, M.Kom dan Lailil Muflikhah, S.Kom., M.Sc selaku Dosen Pembimbing Skripsi.
2. Drs. Marji, M.T. selaku Ketua Program Studi Teknik Informatika dan Ilmu Komputer Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.
3. Putra Pandu Adikara, S.Kom. selaku Dosen Penasehat Akademik.
4. Ir. Sutrisno, M.T. selaku Ketua Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.
5. Segenap Bapak dan Ibu dosen yang telah mendidik dan mengajarkan ilmunya kepada penulis selama menempuh pendidikan di Fakultas MIPA dan Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.
6. Segenap staf dan karyawan di Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya yang telah banyak membantu penulis dalam pelaksanaan penyusunan skripsi ini.
7. Kedua orang tua penulis, Murli Wardesri B.Sc dan Ir. Salmi yang tidak pernah berhenti dalam memberikan do'a, cinta, kasih sayang serta dukungan kepada penulis selama ini.
8. Kedua kakak penulis, Hendri dan Donna, yang selalu memberi dukungan dan motivasi agar selesainya skripsi ini.

9. Keluarga besar penulis, yang selalu memberikan dukungan dan do'a kepada penulis.
10. Sahabat penulis yang telah memberi dukungan baik dalam bentuk material maupun non material demi terselesaikan skripsi ini.
11. Teman-teman Ilmu Komputer angkatan 2009 yang telah banyak memberikan bantuan dan pengalaman selama menjadi mahasiswa di Universitas Brawijaya.
12. Ikatan Pelajar Pemuda Mahasiswa Bundo Kandung (IPPMBK) Malang yang memberikan dukungan dan menjadi keluarga selama penulis menimba ilmu di kota Malang.
13. Penghuni Kos Bunga Anyelir no. 12 atas doa dan motivasinya.
14. Semua pihak yang telah membantu terselesaikannya skripsi ini yang tidak dapat penulis sebutkan satu per satu.

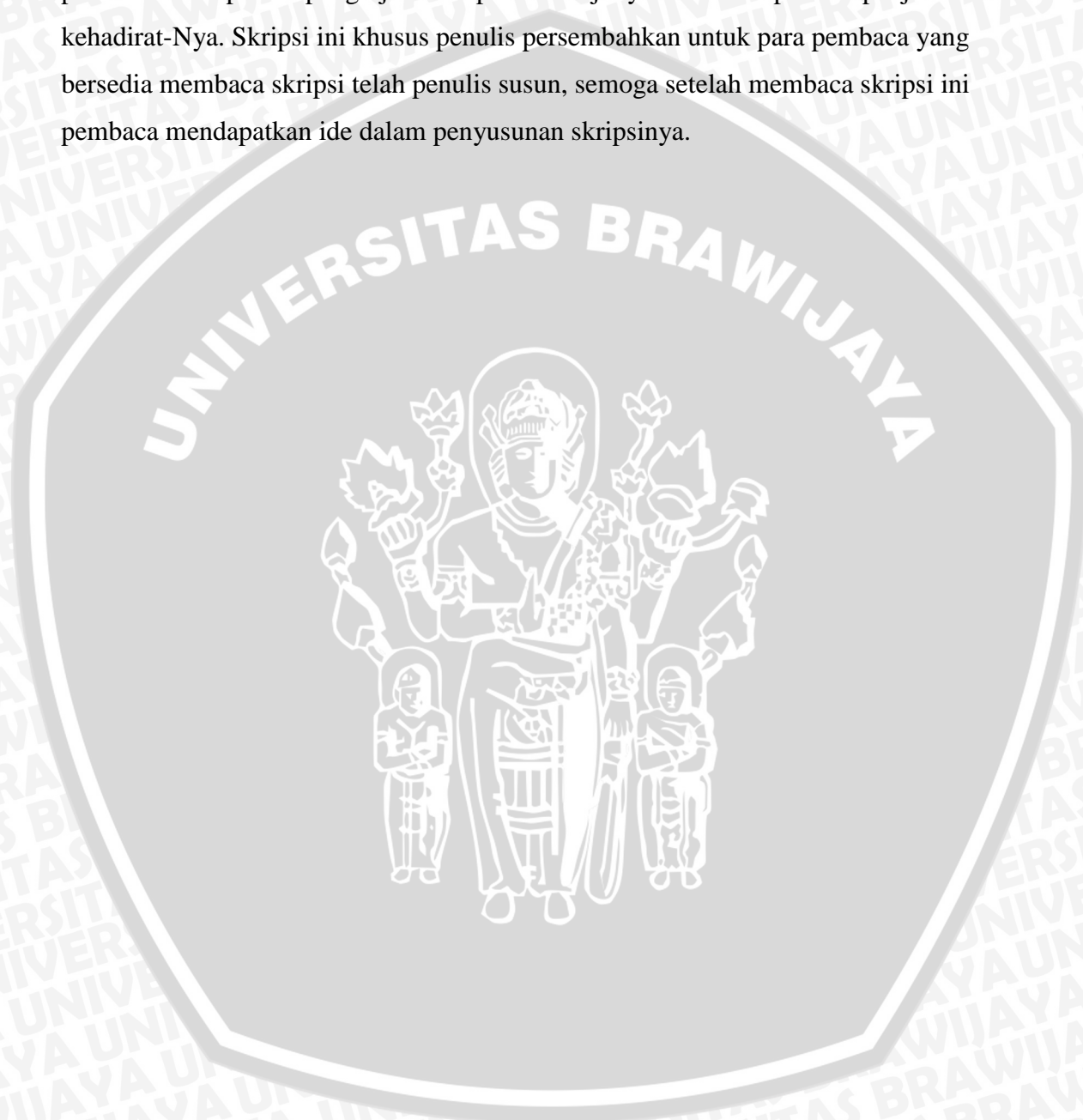
Penulis menyadari bahwa skripsi ini masih banyak kekurangan dan jauh dari sempurna, karena keterbatasan materi dan pengetahuan yang dimiliki penulis. Maka, saran dan kritik yang membangun dari semua pihak sangat diharapkan demi penyempurnaan selanjutnya. Semoga skripsi ini dapat bermanfaat dan berguna bagi semua pihak, baik penulis maupun pembaca, dan semoga Allah SWT meridhoi dan dicatat sebagai ibadah. Amin.

Malang, 28 November 2013

Penulis

## HALAMAN PERSEMBAHAN

Alhamdulillah, terima kasih kepada Allah SWT yang telah memudahkan penulis dalam proses pengerjaan skripsi ini. Puji syukur selalu penulis panjatkan kehadirat-Nya. Skripsi ini khusus penulis persembahkan untuk para pembaca yang bersedia membaca skripsi telah penulis susun, semoga setelah membaca skripsi ini pembaca mendapatkan ide dalam penyusunan skripsinya.



## ABSTRAK

**Yanita Selly Meristika. 2013. Perbandingan K-Nearest Neighbor dan Fuzzy K-Nearest Neighbor pada Diagnosis Penyakit Diabetes Melitus. Skripsi Program Studi Teknik Informatika/Ilmu Komputer, Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya. Pembimbing : Drs. Achmad Ridhok, M.Kom dan Lailil Muflikhah, S.Kom.**

Pada perkembangan di dunia kedokteran saat ini, peneliti dan praktisi memusatkan perhatiannya untuk mendeteksi Diabetes Melitus (DM) dan mencegah atau menghambat berkembangnya komplikasi. Hal ini dikarenakan banyaknya pasien terdiagnosis DM setelah terjadi komplikasi. Padahal DM bisa diatasi jika dideteksi lebih cepat. Salah satu metode untuk pendeteksiannya dapat menggunakan teknik data mining. Pada penelitian ini dilakukan perbandingan antara dua metode yaitu *K-Nearest Neighbor* (K-NN) dan *Fuzzy K-Nearest Neighbor* (FK-NN) untuk mendeteksi DM. *Dataset* DM diambil dari repositori UCI diabetes Indian Pima yang terdiri dari data klinis pasien terdeteksi positif dan negatif DM. K-NN merupakan teknik klasifikasi yang melakukan prediksi secara tegas pada data uji berdasarkan k tetangg terdekat. Sedangkan FK-NN melakukan prediksi data uji menggunakan nilai keanggotaan pada data uji di tiap kelas, kemudian diambil kelas dengan nilai keanggotaan terbesar dari data uji sebagai kelas hasil prediksi. Pengujian ini dilakukan terhadap 4 jumlah data latih yang berbeda yaitu 80, 130, 180, dan 230 dengan menggunakan jumlah data uji yang sama yaitu 50 data. Hasil pengujian yang dilakukan menunjukkan tingkat akurasi tertinggi terdapat pada FK-NN yakni mencapai 98%. Sedangkan K-NN akurasi tertingginya hanya mencapai 96%. Ini berarti *Fuzzy K-Nearest Neighbor* memberikan prediksi yang lebih baik dibandingkan *K-Nearest Neighbor*.

**Kata kunci :** Diabetes Melitus, Indian Pima, perbandingan, FKNN, KNN

## ABSTRACT

**Yanita Selly Meristika. 2013. Comparison of K-Nearest Neighbor and Fuzzy K-Nearest Neighbor on the Diagnosis of Diabetes Mellitus Disease. Minor Thesis Program of Study Information Technology / Computer Science, Program of Technology Information and Computer Science University of Brawijaya. Advisor : Drs. Achmad Ridhok, M.Kom and Lailil Muflikhah, S.Kom.**

*On the progress of medical world today, researchers and practitioners are focusing on detect of diabetes mellitus (DM) and prevent or inhibit the complication evolving. It is because of there are many patients diagnosed with diabetes after complications occur. Whereas, DM can be treated if detected more quickly. One of the methods to detect them is data mining techniques. In this research, I use the comparison between the two methods to detect diabetes, they are the K-Nearest Neighbor (K-NN) and Fuzzy K-Nearest Neighbor (FK-NN). DM datasets are taken from the UCI repository diabetic of Pima Indians that consists of clinical data patients who are diagnosed with positive and negative diabetes. K-NN is a classification technique that explicitly predict test data based on the nearest neighbors of k. While FK-NN predict data using the membership values on test data in each clas, and then taken a class with the largest membership value of the test data as a class prediction results. This Research is conducted on 4 different amount of training data, they are 80, 130, 180, and 230 then each of them using 50 data. The results of the tests show the highest level of accuracy is found in the FK - NN, reaching 98%. While K - NN highest accuracy only reaches 96%. It means that Fuzzy K-Nearest Neighbor gives a better prediction than the K - Nearest Neighbor.*

**Keywords :** Diabetes Mellitus, Pima Indians, comparison, FKNN, KNN

**DAFTAR ISI**

<b>KATA PENGANTAR</b> .....	i
<b>HALAMAN PERSEMBAHAN</b> .....	iii
<b>ABSTRAK</b> .....	iv
<b>DAFTAR ISI</b> .....	vi
<b>DAFTAR TABEL</b> .....	ix
<b>DAFTAR GAMBAR</b> .....	x
<b>DAFTAR LAMPIRAN</b> .....	xiii
<b>DAFTAR SOURCE CODE</b> .....	xiv
<b>BAB I PENDAHULUAN</b> .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	3
1.3. Batasan Masalah .....	3
1.4. Tujuan .....	4
1.5. Manfaat .....	4
1.6. Sistematika Penulisan .....	4
<b>BAB II TINJAUAN PUSTAKA</b> .....	6
2.1. Kajian Pustaka .....	6
2.2. Dasar Teori .....	7
2.2.1. Diabetes Melitus .....	7
2.2.1.1. Pengertian Diabetes Melitus .....	7
2.2.1.2. Data Diabetes Indian Pima .....	8
2.2.2. Logika Fuzzy .....	9
2.2.2.1. Pengertian Logika Fuzzy .....	9
2.2.2.2. Himpunan Fuzzy .....	9
2.2.3. Data Mining .....	11
2.2.3.1. Pengertian Data Mining .....	11
2.2.3.2. Klasifikasi .....	12
2.2.4. K-Nearest Neighbor .....	13
2.2.4.1. Proses K-Nearest Neighbor .....	14
2.2.5. Fuzzy K-Nearest Neighbor .....	15

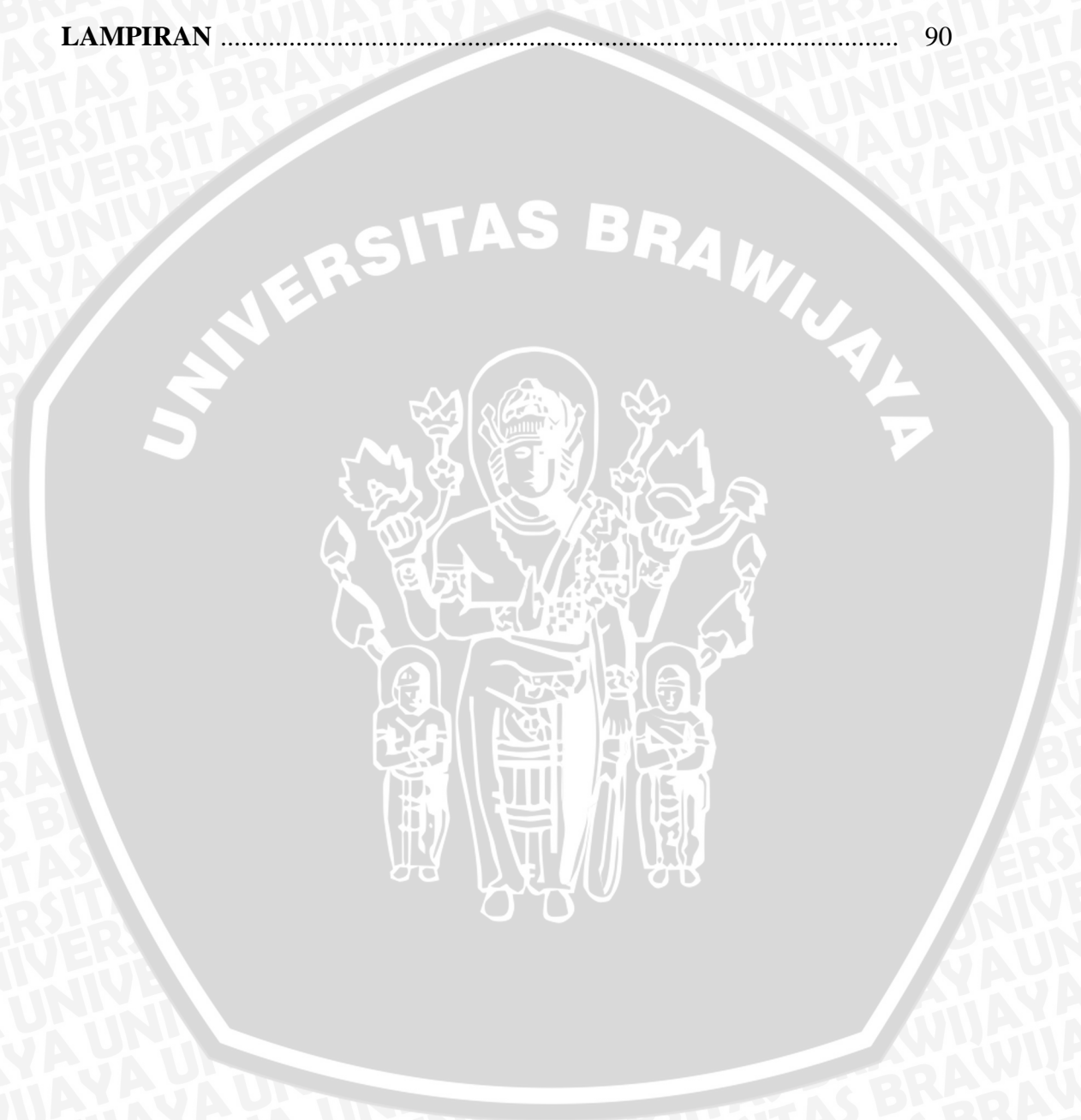


2.2.5.1. Definisi Fuzzy K-Nearest Neighbor .....	15
2.2.5.2. Algoritma Fuzzy K-Nearest Neighbor .....	16
2.2.6. Akurasi .....	17
<b>BAB III METODOLOGI DAN PERANCANGAN .....</b>	<b>19</b>
3.1. Studi Literatur .....	20
3.2. Menentukan dan Mempelajari Dataset yang digunakan .....	20
3.3. Analisa dan Perancangan Sistem .....	21
3.3.1. Proses Klasifikasi K-Nearest Neighbor dan Fuzzy K-Nearest Neighbor .....	22
3.3.2. Flowchart Sistem .....	23
3.3.3. Normalisasi Nilai Atribut .....	24
3.3.4. Klasifikasi Menggunakan K-Nearest Neighbor .....	25
3.3.5. Klasifikasi Menggunakan Fuzzy K-Nearest Neighbor .....	26
3.4. Perhitungan Manual .....	30
3.4.1. Normalisasi Nilai Atribut .....	30
3.4.2. Menghitung Jarak Record Baru Data Uji di Setiap Record Data Latih .....	32
3.4.3. Menentukan K Record Terdekat .....	33
3.4.4. K-Nearest Neighbor (K-NN) .....	33
3.4.5. Proses Fuzzy K-Nearest Neighbor (FK-NN) .....	33
3.5. Perancangan Antarmuka .....	35
3.5.1. Form Prediksi .....	35
3.5.2. Form Koleksi Data Uji .....	36
3.5.3. Perbandingan Akurasi .....	40
3.6. Perancangan Uji Coba .....	41
3.6.1. Uji Pengaruh Nilai k dan Data Latih Terhadap Tingkat Akurasi .....	42
3.6.2. Uji Pengaruh Nilai m Terhadap Tingkat Akurasi .....	43
<b>BAB IV IMPLEMENTASI .....</b>	<b>44</b>
4.1. Lingkungan Implementasi .....	44
4.1.1. Lingkungan Implementasi Perangkat Keras .....	44
4.1.2. Lingkungan Implementasi Perangkat Lunak .....	44

4. 2. Implementasi Program .....	44
4.2.1. Implementasi Class .....	45
4.2.2. Tahap Proses Pelatihan .....	51
4.2.2.1. Proses normalisasi nilai setiap atribut .....	52
4.2.2.2. Proses klasifikasi menggunakan algoritma k-Nearest Neighbor .....	53
4.2.2.3. Proses klasifikasi menggunakan algoritma fuzzy k-Nearest Neighbor .....	56
4. 3. Implementasi Antarmuka .....	58
4.3.1. Form Prediksi .....	58
4.3.2. Form Koleksi Data Uji .....	60
4.3.3. Form Perbandingan Akurasi .....	69
<b>BAB V PENGUJIAN DAN ANALISIS</b> .....	72
5. 1. Pengujian Sistem.....	72
5.1.1. Sistematika pengujian pengaruh nilai k dan data latih .....	72
5.1.2. Sistematika pengujian nilai m terhadap tingkat akurasi .....	72
5. 2. Hasil Pengujian .....	73
5.2.1. Hasil uji pengaruh nilai k dan data latih .....	73
4.2.2.4. Pengujian pengaruh jumlah data latih .....	73
4.2.2.5. Pengujian pada jumlah data latih 80 .....	74
4.2.2.6. Pengujian pada jumlah data latih 130 .....	74
4.2.2.7. Pengujian pada jumlah data latih 180 .....	75
4.2.2.8. Pengujian pada jumlah data latih 230 .....	76
5.2.2. Hasil uji pengaruh dataset kelas seimbang ( <i>balanced class</i> ) dan tidak seimbang ( <i>imbalanced class</i> ) .....	77
5.2.3. Hasil uji pengaruh nilai bobot .....	78
5. 3. Analisa Hasil Uji Coba .....	78
5.3.1. Analisa pengaruh jumlah data latih .....	78
5.3.2. Analisa pengaruh nilai k .....	80
5.3.3. Analisa pengaruh dataset kelas seimbang ( <i>balanced class</i> ) dan tidak seimbang ( <i>imbalanced class</i> ) .....	82
5.3.4. Pengaruh nilai bobot pada FK-NN .....	83



<b>BAB VI PENUTUP</b> .....	86
6. 1. Kesimpulan .....	86
6. 2. Saran .....	87
<b>DAFTAR PUSTAKA</b> .....	88
<b>LAMPIRAN</b> .....	90



DAFTAR TABEL

Tabel 2.1	Penggunaan data diabetes Indian Pima dengan metode lainnya ..	6
Tabel 2.2	Penelitian menggunakan K-NN dan FK-NN .....	6
Tabel 3.1	Jumlah <i>missing value</i> pada setiap attribut .....	20
Tabel 3.2	Contoh data latih sistem .....	30
Tabel 3.3	Contoh data uji sistem .....	30
Tabel 3.4	Range data .....	31
Tabel 3.5	Data latih yang telah dinormalisasi .....	31
Tabel 3.6	Data uji yang telah dinormalisasi .....	32
Tabel 3.7	Hasil perhitungan jarak dan diurutkan dari yang terkecil .....	32
Tabel 3.8	Tabel Perancangan Uji Coba.....	42
Tabel 3.9	Tabel uji pengaruh nilai k dan data latih terhadap akurasi sistem	43
Tabel 3.10	Tabel uji pengaruh nilai m (bobot) terhadap tingkat akurasi .....	43
Tabel 4.1	Deskripsi metode/fungsi pada <i>package</i> K-NN.....	50
Tabel 4.2	Deskripsi metode/fungsi pada <i>package</i> FK-NN.....	51
Tabel 5.1	Pengaruh jumlah data latih pada KNN dan FKNN .....	73
Tabel 5.2	Hasil pengujian jenis <i>dataset</i> terhadap KNN .....	77
Tabel 5.3	Hasil pengujian jenis <i>dataset</i> terhadap FKNN.....	78
Tabel 5.4	Tabel Hasil Uji Coba Pengaruh Bobot Pada FK-NN .....	78

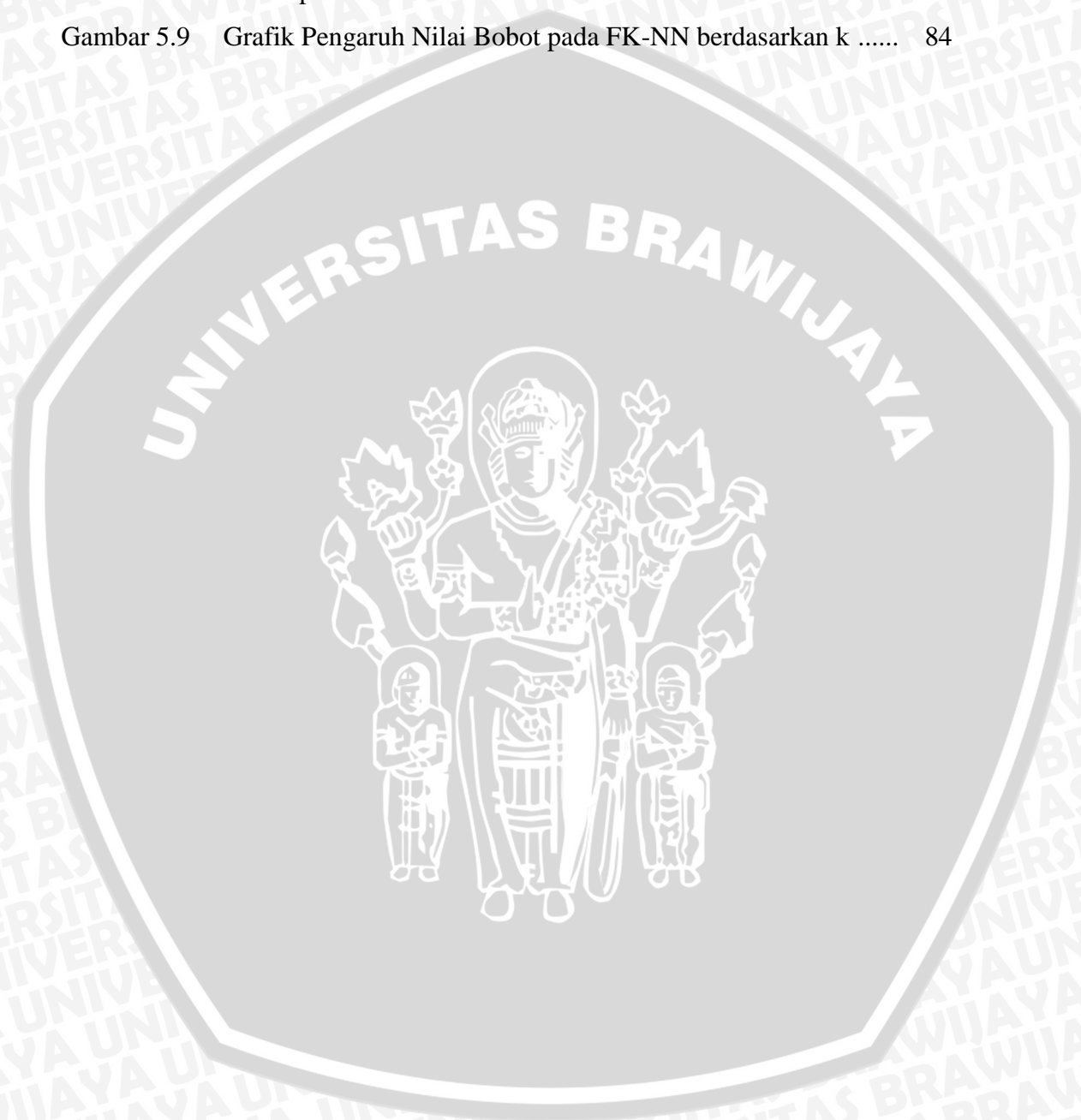


**DAFTAR GAMBAR**

Gambar 3.1	Diagram Alir Penelitian.....	19
Gambar 3.2	Alur proses sistem .....	23
Gambar 3.3	Alur proses Normalisasi atribut .....	25
Gambar 3.4	Alur proses K-NN .....	27
Gambar 3.5	Proses <i>Euclidean Distance</i> .....	28
Gambar 3.6	Alur Proses FK-NN.....	29
Gambar 3.7a	<i>Interface</i> form prediksi K-NN .....	35
Gambar 3.7b	<i>Interface</i> form prediksi FK-NN .....	36
Gambar 3.8a	<i>Interface</i> koleksi data uji K-NN dan FK-NN .....	37
Gambar 3.8b	<i>Interface</i> koleksi data uji tab proses FK-NN.....	38
Gambar 3.8c	<i>Interface</i> koleksi data uji tab proses K-NN .....	38
Gambar 3.9	Form perbandingan akurasi K-NN dan F-KNN .....	40
Gambar 4.1	Antarmuka form prediksi K-NN .....	59
Gambar 4.2	Antarmuka form prediksi FK-NN .....	60
Gambar 4.3	Antarmuka pada tab input data K-NN .....	61
Gambar 4.4	Antarmuka pada tab proses K-NN .....	62
Gambar 4.5	Antarmuka tab K pada K-NN .....	63
Gambar 4.6	Antarmuka tab akurasi pada K-NN .....	64
Gambar 4.7	Antarmuka input data FK-NN .....	65
Gambar 4.8	Antarmuka proses FK-NN .....	66
Gambar 4.9	Antarmuka tab k pada FK-NN .....	67
Gambar 4.10	Antarmuka tab membership pada FK-NN .....	68
Gambar 4.11	Antarmuka input data perbandingan akurasi .....	70
Gambar 4.12	Antarmuka data perbandingan akurasi K-NN dan FK-NN .....	71
Gambar 5.1	Grafik pengaruh nilai k pada 80 data latih .....	74
Gambar 5.2	Grafik pengaruh nilai k pada 130 data latih .....	75
Gambar 5.3	Grafik pengaruh nilai k pada 180 data latih .....	77
Gambar 5.4	Grafik pengaruh nilai k pada 230 data latih .....	74
Gambar 5.5	Pengaruh Jumlah Data Latih pada K-NN .....	79
Gambar 5.6	Pengaruh Jumlah Data Latih pada FK-NN .....	80

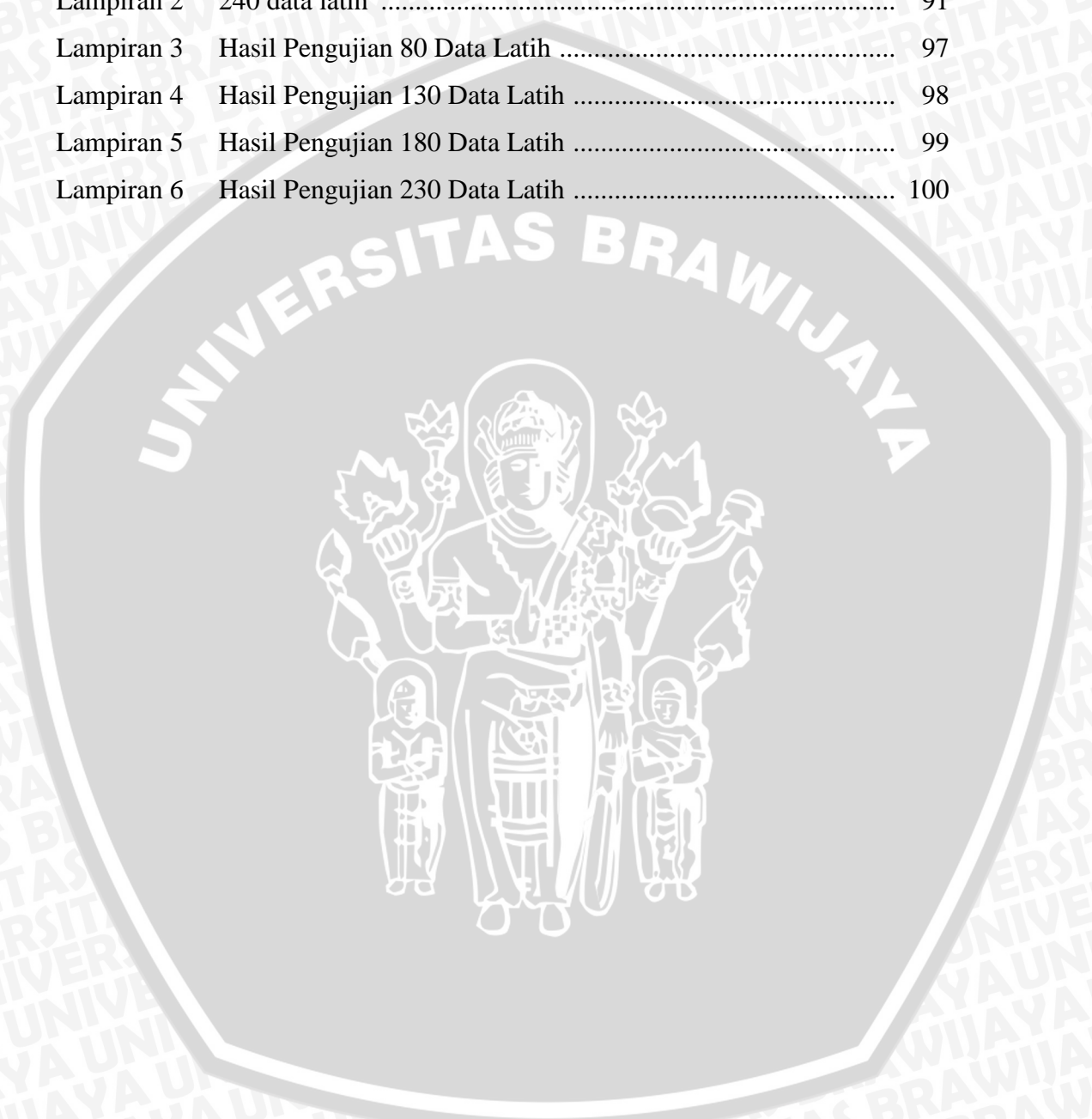


Gambar 5.7	Grafik Pengaruh Keseimbangan Jumlah Data dalam Setiap Kelas pada K-NN .....	82
Gambar 5.8	Grafik Pengaruh Keseimbangan Jumlah Data dalam Setiap Kelas pada FK-NN .....	83
Gambar 5.9	Grafik Pengaruh Nilai Bobot pada FK-NN berdasarkan k .....	84



**DAFTAR LAMPIRAN**

Lampiran 1	50 data uji .....	90
Lampiran 2	240 data latih .....	91
Lampiran 3	Hasil Pengujian 80 Data Latih .....	97
Lampiran 4	Hasil Pengujian 130 Data Latih .....	98
Lampiran 5	Hasil Pengujian 180 Data Latih .....	99
Lampiran 6	Hasil Pengujian 230 Data Latih .....	100



## DAFTAR SOURCE CODE

Sourcecode 4.1 Sourcecode penyimpanan data .....	45
Sourcecode 4.2 Sourcecode cetak data .....	46
Sourcecode 4.3 Sourcecode kelas dataK-NN .....	47
Sourcecode 4.4 Sourcecode Penyimpanan variabel <i>membership</i> .....	47
Sourcecode 4.5 Sourcecode pembacaan file excel .....	50
Sourcecode 4.6 Normalisasi nilai atribut data latih dan data uji .....	53
Sourcecode 4.7 Fungsi menghitung jarak menggunakan <i>euclidean</i> .....	54
Sourcecode 4.8 Proses pengambilan nilai k berdasarkan jarak terkecil menggunakan <i>buble sort</i> .....	55
Sourcecode 4.9 Perhitungan jumlah data training berdasarkan kelas .....	55
Sourcecode 4.10 Pencarian kelas mayoritas sebagai hasil akhir klasifikasi ...	56
Sourcecode 4.11 Perhitungan nilai derajat keanggotaan .....	57
Sourcecode 4.12 Cari nilai <i>membership</i> terbesar .....	58

