

**KETEPATAN KLASIFIKASI STATUS KERJA
DI KOTA TEGAL MENGGUNAKAN ALGORITMA C4.5 DAN
*FUZZY K-NEAREST NEIGHBOR IN EVERY CLASS (FK-NNC)***



SKRIPSI

Disusun Oleh :
ATIKA ELSADINING TYAS
24010211120013

JURUSAN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG

2015

**KETEPATAN KLASIFIKASI STATUS KERJA
DI KOTA TEGAL MENGGUNAKAN ALGORITMA C4.5 DAN
*FUZZY K-NEAREST NEIGHBOR IN EVERY CLASS (FK-NNC)***

Disusun Oleh:

ATIKA ELSADINING TYAS

24010211120013

Skripsi

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Sains pada
Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro

**JURUSAN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG
2015**

HALAMAN PENGESAHAN I

Judul : Ketepatan Klasifikasi Status Kerja di Kota Tegal Menggunakan
Algoritma C4.5 dan *Fuzzy K-Nearest Neighbor in every Class*
(FK-NNC)

Nama : Atika Elsadining Tyas

NIM : 24010211120013

Telah diujikan pada sidang Tugas Akhir tanggal 31 Agustus 2015 dan dinyatakan
lulus pada tanggal 10 September 2015

Semarang, September 2015

Mengetahui,

Ketua Jurusan Statistika

FSM UNDIP



Drs. H. Dwi Ispriyanti, M.Si

NIP. 195709141986032001

Panitia Penguji Ujian Tugas Akhir

Ketua,

A handwritten signature in black ink, appearing to be 'S. Suparti', written over a faint grid background.

Dr. Suparti, M.Si

NIP. 196509131990032001

HALAMAN PENGESAHAN II

Judul : Ketepatan Klasifikasi Status Kerja di Kota Tegal Menggunakan
Algoritma C4.5 dan *Fuzzy K-Nearest Neighbor in every Class*
(FK-NNC)

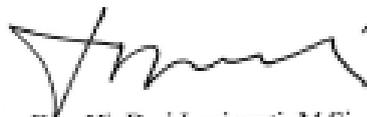
Nama : Atika Elsadining Tyas

NIM : 24010211120013

Telah diujikan pada sidang Tugas Akhir tanggal 31 Agustus 2015.

Semarang, September 2015

Dosen Pembimbing I



Dra. Hj. Dwi Ispriyanti, M.Si

NIP. 195709141986032001

Dosen Pembimbing II



Drs. Sudarno, M.Si

NIP. 196407091992011001

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT atas rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang diberi judul “**Ketepatan Klasifikasi Status Kerja di Kota Tegal Menggunakan Algoritma C4.5 dan *Fuzzy K-Nearest Neighbor in every Class (FK-NNC)***”. Tugas Akhir ini tidak akan terselesaikan dengan baik tanpa adanya dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Ibu Dra. Hj. Dwi Ispriyanti, M.Si. selaku Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro dan dosen pembimbing I.
2. Bapak Drs. Sudarno, M.Si selaku dosen pembimbing II.
3. Bapak/Ibu dosen jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
4. Semua pihak yang telah membantu kelancaran penyusunan Tugas Akhir ini, yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran demi kesempurnaan penulisan selanjutnya.

Semarang, September 2015

Penulis

ABSTRAK

Masalah pengangguran merupakan masalah yang sangat krusial yang selalu dihadapi negara berkembang dan dapat mempengaruhi pembangunan nasional. Digunakan dua metode untuk melakukan klasifikasi status kerja pada penduduk angkatan kerja di Kota Tegal bulan Agustus tahun 2014 yaitu algoritma C4.5 dan *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC). Algoritma C4.5 adalah salah satu metode klasifikasi dari *data mining* yang digunakan untuk mengkonstruksikan pohon keputusan (*decision tree*). *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC) adalah teknik klasifikasi yang melakukan prediksi menggunakan sejumlah K tetangga terdekat pada setiap kelas dari sebuah data uji. Variabel prediktor yang digunakan adalah status dalam rumah tangga, jenis kelamin, umur, status perkawinan, pendidikan, dan pelatihan kerja. Untuk mengevaluasi hasil klasifikasi digunakan perhitungan APER. Berdasarkan analisis tersebut pengklasifikasian status kerja dengan Algoritma C4.5 diperoleh nilai APER = 28,3784% dan ketepatan klasifikasi sebesar 71,6216% sedangkan menggunakan metode FK-NNC diperoleh nilai APER = 21,62% dan ketepatan klasifikasi sebesar 78,38%. Sehingga dapat disimpulkan bahwa metode FK-NNC lebih baik dibandingkan dengan Algoritma C4.5.

Kata kunci: Klasifikasi, Algoritma C4.5, *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC), APER

ABSTRACT

Unemployment is a very crucial problem that always deal a developing country and affected a national foundation. It used two methods for classifying a employment status on productive society in Tegal City on August 2014, the methods are C4.5 Algorithm and *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC). C4.5 Algorithm is a way of classifying methods from data mining that use to construct a decision tree. FK-NNC is another classification technique that predict using the amount of closest neighbor of K in every class from a testing data. The predictor variables that used on classifying an employment status are neighborhood status, sex, age, marriage status, education, and a work training. To evaluate the result of classification use APER calculation. Based on this analysis, classification of employment status using C4.5 Algorithm obtained APER = 28,3784% and 71,6216% of accuracy, while FK-NNC methods obtained APER = 21,62% and 78,38% of accuracy. So, it can be concluded that FK-NNC is better than C4.5 Algorithm.

Keywords: Classification, C4.5 Algorithm, *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC), APER

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN I	ii
HALAMAN PENGESAHAN II	iii
KATA PENGANTAR	iv
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan	4
BAB II TINJAUAN PUSTAKA	
2.1 Ketenagakerjaan	6
2.2 <i>Data Mining</i>	12
2.3 Algoritma C4.5	13
2.3.1 Pembentukan Pohon Keputusan Algoritma C4.5	15
2.3.2 Prosedur Pemilahan Algoritma C4.5	17
2.3.3 Contoh Pemilahan Pohon Keputusan Algoritma C4.5 ...	18

2.4	<i>Fuzzy K-Nearest Neighbor in every Class</i>	22
2.5	Ketepatan Klasifikasi	27
BAB III METODOLOGI PENELITIAN		
3.1	Jenis dan Sumber Data	29
3.2	Variabel Penelitian	29
3.3	Teknik Pengolahan Data	30
3.4	Tahapan Penelitian	31
BAB IV HASIL DAN PEMBAHASAN		
4.1	Analisis Deskriptif Data Status Kerja	33
4.1.1	Status Kerja Kota Tegal Bulan Agustus 2014.....	33
4.1.2	Status Kerja Kota Tegal Bulan Agustus 2014 Berdasarkan Status dalam Rumah Tangga	34
4.1.3	Status Kerja Kota Tegal Bulan Agustus 2014 Berdasarkan Jenis Kelamin	36
4.1.4	Status Kerja Kota Tegal Bulan Agustus 2014 Berdasarkan Umur	37
4.1.5	Status Kerja Kota Tegal Bulan Agustus 2014 Berdasarkan Status Perkawinan	37
4.1.6	Status Kerja Kota Tegal Bulan Agustus 2014 Berdasarkan Pendidikan	39
4.1.7	Status Kerja Kota Tegal Bulan Agustus 2014 Berdasarkan Pelatihan Kerja	40
4.2	Pengklasifikasian Menggunakan Algoritma C4.5	41
4.2.1	Konstruksi Algoritma C4.5	42

4.2.2 Analisis Pohon Keputusan	47
4.2.3 Identifikasi Status Kerja	48
4.2.4 Pengujian Hasil Pohon Keputusan	50
4.3 Pengklasifikasian Menggunakan Metode FK-NNC	50
4.4 Perbandingan Ketepatan Klasifikasi	57
BAB V KESIMPULAN	58
DAFTAR PUSTAKA	59
LAMPIRAN	61

DAFTAR TABEL

	Halaman
Tabel 1	Data Klasifikasi Bermain <i>Baseball</i> dengan Tipe Atribut Campuran . 19
Tabel 2	Hasil Perhitungan Nilai v untuk Atribut Suhu..... 20
Tabel 3	Hasil Perhitungan <i>Entropy</i> dan <i>Gain</i> untuk <i>Node</i> Akar 21
Tabel 4	Formula Jarak Dua Data dengan Satu Atribut..... 23
Tabel 5	Hasil Perhitungan Jarak FK-NNC pada Set Data Buatan..... 25
Tabel 6	Matriks Konfusi..... 27
Tabel 7	Variabel Penelitian 29
Tabel 8	Status Kerja Kota Tegal Bulan Agustus 2014 34
Tabel 9	Status Kerja Berdasarkan Status dalam Rumah Tangga 35
Tabel 10	Status Kerja Berdasarkan Jenis Kelamin 36
Tabel 11	Deskriptif untuk Variabel Umur 37
Tabel 12	Status Kerja Berdasarkan Status Perkawinan 38
Tabel 13	Status Kerja Berdasarkan Pendidikan 39
Tabel 14	Status Kerja Berdasarkan Pelatihan Kerja 40
Tabel 15	Frekuensi Tiap Kelas 42
Tabel 16	Peluang Masing-masing Kelas 42
Tabel 17	Frekuensi Masing-masing Kategori pada Atribut Jenis Kelamin Berdasarkan Kelasnya 43
Tabel 18	Peluang Masing-masing Kategori Berdasarkan Kelasnya 43
Tabel 19	Nilai <i>Information Gain</i> pada Simpul Akar 44

Tabel 20	Nilai Ambang Batas Atribut Umur pada Proses Pembentukan Pohon Keputusan Tingkat ke-empat	46
Tabel 21	Matriks Konfusi Data <i>Testing</i>	50
Tabel 22	Data <i>Training</i>	52
Tabel 23	Data <i>Testing</i>	52
Tabel 24	Ketepatan Klasifikasi FK-NNC untuk $K = 3$	56
Tabel 25	Hasil Ketepatan Klasifikasi Menggunakan FK-NNC	56

DAFTAR GAMBAR

	Halaman
Gambar 1 Contoh Pohon Keputusan (<i>Decision Tree</i>)	14
Gambar 2 Konsep K Tetangga Terdekat dari Setiap Kelas dalam F-KNNC	22
Gambar 3 Diagram Alir Pengolahan Data Menggunakan Algoritma C4.5 dan F-KNNC	32
Gambar 4 Diagram Status Kerja	34
Gambar 5 Diagram Status Kerja Berdasarkan Status dalam Rumah Tangga	35
Gambar 6 Diagram Status Kerja Berdasarkan Jenis Kelamin	36
Gambar 7 Diagram Status Kerja Berdasarkan Status Perkawinan	38
Gambar 8 Diagram Status Kerja Berdasarkan Pendidikan	39
Gambar 9 Diagram Status Kerja Berdasarkan Pelatihan Kerja	41
Gambar 10 Pohon Keputusan Tingkat Pertama	45

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Data SAKERNAS Kota Tegal Bulan Agustus Tahun 2014	61
Lampiran 2 Hasil Algoritma C4.5 Menggunakan Data <i>Training</i>	62
Lampiran 3 Pohon Keputusan yang Terbentuk Menggunakan Data <i>Training</i>	65
Lampiran 4 Hasil Algoritma C4.5 Menggunakan Data <i>Testing</i>	66
Lampiran 5 Hasil Pengklasifikasian Menggunakan Algoritma C4.5.....	67
Lampiran 6 <i>Output</i> Ketepatan Klasifikasi Menggunakan FK-NNC Untuk $K = 3$	73
Lampiran 7 Syntax Matlab untuk FK-NNC	77

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan negara kepulauan yang memiliki jumlah penduduk yang besar, serta dikategorikan sebagai negara berkembang. Suatu negara dikatakan berkembang atau maju salah satunya adalah dengan melihat pada keberhasilan pembangunan oleh negara yang bersangkutan. Indonesia sebagai negara berkembang perlu melakukan banyak perubahan untuk mendukung pembangunan nasional. Pembangunan nasional bertujuan untuk meningkatkan kesejahteraan masyarakat. Meningkatnya kesejahteraan rakyat yang adil dan merata merupakan tujuan dari pembangunan nasional Indonesia. Namun, dimensi kesejahteraan rakyat disadari sangat luas dan kompleks. Suatu taraf kesejahteraan rakyat hanya dapat dinilai melalui indikator-indikator yang terukur dari berbagai aspek pembangunan. Menurut BPS (2014), aspek-aspek yang dapat menggambarkan kondisi tingkat kesejahteraan rakyat pada saat ini meliputi kependudukan, kesehatan dan gizi, pendidikan, ketenagakerjaan, taraf dan pola konsumsi, perumahan dan lingkungan, kemiskinan, serta sosial lainnya.

Salah satu masalah yang menjadi perhatian pemerintah adalah permasalahan di bidang ketenagakerjaan terutama masalah pengangguran. Masalah pengangguran merupakan masalah yang sangat krusial yang selalu dihadapi negara berkembang dan dapat mempengaruhi pembangunan nasional. Tingginya pengangguran di suatu negara mengakibatkan rendahnya partisipasi rakyat dalam pertumbuhan ekonomi yaitu pendapatan rakyat akan berkurang

sehingga dapat menyebabkan timbulnya kemiskinan dan masalah sosial lainnya. Dari data Badan Pusat Statistik (BPS) diketahui angkatan kerja Indonesia pada Agustus 2014 mencapai angka 121,9 juta orang. Sedangkan, penduduk yang bekerja pada Agustus 2014 adalah sebanyak 114,6 juta orang. Untuk angka Tingkat Pengangguran Terbuka (TPT) pada Agustus 2014, menurut Kepala BPS Suryamin adalah sebesar 5,94%. Jika dibandingkan dengan Agustus 2013, tingkat pengangguran terbuka mengalami penurunan dari 6,17% menjadi 5,94%. Berdasarkan permasalahan di atas, perlu dilakukan pengklasifikasian status kerja bahwa seseorang dikatakan pengangguran atau bukan pengangguran.

Pada tugas akhir ini digunakan dua metode untuk mengklasifikasikan data status kerja tersebut, yaitu Algoritma C4.5 dan *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC), dimana kedua metode tersebut merupakan teknik klasifikasi yang sederhana tetapi mempunyai hasil kerja yang cukup bagus. Algoritma C4.5 adalah salah satu metode klasifikasi dari *data mining* yang digunakan untuk mengkonstruksikan pohon keputusan (*decision tree*). Menurut Prasetyo (2014), pohon keputusan atau *decision tree* adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Algoritma C4.5 diperkenalkan oleh Quinlan (1993) sebagai versi perbaikan dari algoritma *Iterative Dichotomiser 3* (ID3). Menurut Witten *et al.*, (2011), Algoritma C4.5 memiliki keunggulan dibandingkan dengan ID3 yaitu mampu mengatasi nilai yang hilang (*missing value*), mengatasi data bertipe kontinu, dan melakukan pemangkasan pohon (*prunning trees*). Selain itu, dengan menggunakan Algoritma C4.5 dapat diketahui pula nilai ketepatan klasifikasi.

Fuzzy K-Nearest Neighbor in every Class atau dapat disingkat FK-NNC adalah teknik klasifikasi yang melakukan prediksi menggunakan sejumlah K tetangga terdekat pada setiap kelas dari sebuah data uji, bukan K tetangga terdekat seperti pada *K-Nearest Neighbor* (K-NN) dan *Fuzzy K-Nearest Neighbor* (FK-NN). Penelitian sebelumnya pernah dilakukan oleh Eko Prasetyo (2012b) pada klasifikasi *data set* iris dan *vertebral column* dengan membandingkan tiga metode yaitu metode *K-Nearest Neighbor* (K-NN), *Fuzzy K-Nearest Neighbor* (FK-NN), dan *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC). Pada penelitian tersebut didapatkan kesimpulan bahwa nilai akurasi yang diberikan oleh FK-NNC terbukti lebih tinggi daripada K-NN dan FK-NN.

Penelitian status kerja sebelumnya pernah dilakukan oleh Riyan Eko Putri (2014) dengan membandingkan dua metode yaitu Naïve Bayes dan *K-Nearest Neighbor* (K-NN). Pada tugas akhir ini akan diteliti mengenai data status kerja Kota Tegal tahun 2014 menggunakan Algoritma C4.5 yang akan dibandingkan dengan menggunakan metode *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC). Dari kedua metode tersebut masing-masing akan menghasilkan nilai ketepatan klasifikasi. Sehingga, dapat dicari ketepatan klasifikasi yang terbaik dari kedua metode tersebut.

1.2 Rumusan Masalah

Permasalahan yang akan dibahas pada tugas akhir ini adalah sebagai berikut:

1. Bagaimana penerapan Algoritma C4.5 dalam pembentukan pohon klasifikasi untuk data status kerja di Kota Tegal?

2. Bagaimana ketepatan klasifikasi data status kerja untuk Kota Tegal tahun 2014 menggunakan Algoritma C4.5?
3. Bagaimana ketepatan klasifikasi data status kerja untuk Kota Tegal tahun 2014 menggunakan metode *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC)?
4. Bagaimana perbandingan ketepatan klasifikasi data status kerja untuk Kota Tegal tahun 2014 menggunakan Algoritma C4.5 dan metode *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC)?

1.3 Batasan Masalah

Permasalahan pada tugas akhir ini dibatasi untuk daerah Kota Tegal, sesuai dengan pendataan yang dilakukan oleh BPS pada Agustus 2014. Pengolahan tersebut diberikan batasan masalah pada pembahasan tentang perbandingan hasil ketepatan klasifikasi menggunakan dua metode, yaitu Algoritma C4.5 dan metode *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC).

1.4 Tujuan

Tujuan dari penulisan tugas akhir ini adalah:

1. Membentuk pohon klasifikasi untuk data status kerja di Kota Tegal menggunakan Algoritma C4.5.
2. Mengukur ketepatan klasifikasi data status kerja untuk Kota Tegal tahun 2014 menggunakan Algoritma C4.5.

3. Mengukur ketepatan klasifikasi data status kerja untuk Kota Tegal tahun 2014 menggunakan metode *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC).
4. Membandingkan hasil ketepatan klasifikasi status kerja untuk Kota Tegal tahun 2014 menggunakan Algoritma C4.5 dan metode *Fuzzy K-Nearest Neighbor in every Class* (FK-NNC).