

**IN
REVIEW**



**KLASIFIKASI ALGORITMA C4.5, NAÏVE BAYES DAN K-NEAREST
NEIGHBOUR UNTUK PENERIMA BANTUAN PROGRAM KELUARGA
HARAPAN (PKH) WILAYAH DKI JAKARTA**

TUGAS AKHIR

MONICA CHELSEA ESTER DONDOKAMBEY
41516010136

**UNIVERSITAS
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2020**



**KLASIFIKASI ALGORITMA C4.5, NAÏVE BAYES DAN K-NEAREST
NEIGHBOUR UNTUK PENERIMA BANTUAN PROGRAM KELUARGA
HARAPAN (PKH) WILAYAH DKI JAKARTA**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

MONICA CHELSEA ESTER DONDOKAMBEY
41516010136

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA

2020

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41516010136

Nama : Monica Chelsea Ester Dondokambey

Judul Tugas Akhir : Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour Untuk Penerima Bantuan Program Keluarga Harapan (PKH) Wilayah DKI Jakarta

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 17 Febuari 2020

Monica Chelsea Est



UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Monica Chelsea Ester Dondokambey
NIM : 41516010136
Judul Tugas Akhir : Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour Untuk Penerima Bantuan Program Keluarga Harapan (PKH) Wilayah DKI Jakarta

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 17 Februari 2020

UNIVERSITAS
MERCU BUANA


Monica Chelsea Es

METERAI
TEMPEL
20
E1375AHF354980259
6000
ENAM RIBU RUPIAH



SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Monica Chelsea Ester Dondokambey
 NIM : 41516010136
 Judul Tugas Akhir : Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour Untuk Penerima Bantuan Program Keluarga Harapan (PKH) Wilayah DKI Jakarta

Menyatakan bahwa Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan <input checked="" type="checkbox"/>
		Jurnal Nasional Terakreditasi	
		Jurnal International Tidak Bereputasi	Diterima <input type="checkbox"/>
		Jurnal International Bereputasi	
Disubmit/dipublikasikan di :		Nama Jurnal : Jurnal Teknologi dan Sistem Komputer	
		ISSN :	
2	Kertas Kerja, Merupakan material hasil penelitian sebagai kelengkapan Artikel Jurnal. Terdiri dari (minimal 4)	Literatur Review	<input checked="" type="checkbox"/>
		Hasil analisa & perancangan aplikasi	<input type="checkbox"/>
		Source code	<input type="checkbox"/>
		Data set	<input checked="" type="checkbox"/>
		Tahapan eksperimen	<input checked="" type="checkbox"/>
		Hasil eksperimen seluruhnya	<input checked="" type="checkbox"/>
		
3	HAKI Disubmit / Terdaftar	HKI	Diajukan <input type="checkbox"/>
		Paten	Tercatat <input type="checkbox"/>
		No & Tanggal Permohonan :	
		No & Tanggal Pencatatan :	

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 17 F



Monica Chelsea Ester D



LEMBAR PERSETUJUAN PENGUJI

NIM : 41516010136
Nama : Monica Chelsea Ester Dondokambey
Judul Tugas Akhir : Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour Untuk Penerima Bantuan Program Keluarga Harapan (PKH) Wilayah DKI Jakarta

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.


Jakarta, 12 Februari 2020



(Diky Firdaus, S.Kom, MM)
Ketua Penguji



(Anis Cherid, SE, MTI)
Anggota Penguji 1



(Sabar Rudiarto, S.Kom, M.Kom)
Anggota Penguji 2

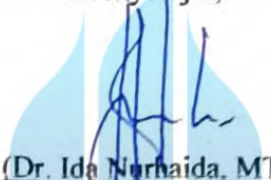
LEMBAR PENGESAHAN

NIM : 41516010136
Nama : Monica Chelsea Ester Dondokambey
Judul Tugas Akhir : Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour Untuk Penerima Bantuan Program Keluarga Harapan (PKH) Wilayah DKI Jakarta

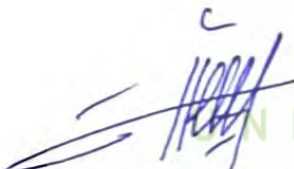
Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.


Jakarta, 17 Februari 2020

Menyetujui,


(Dr. Ida Nurhaida, MT)
Dosen Pembimbing

Mengetahui,


(Diky Firdaus, S.Kom, MM)
Koord. Tugas Akhir Teknik Informatika


(Desi Ramavanti, S.Kom, MT)
Ka. Prodi Teknik Informatika

ABSTRAK

Nama : Monica Chelsea Ester Dondokambey
NIM : 41516010136
Pembimbing TA : Dr. Ida Nurhaida, MT
Judul : Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour Untuk Penerima Bantuan Program Keluarga Harapan (PKH) Wilayah DKI Jakarta

Kemiskinan dan ketimpangan merupakan salah satu permasalahan yang terjadi di beberapa negara. Indonesia merupakan salah satu negara di Asia Tenggara yang memiliki permasalahan kemiskinan tersebut. Program Keluarga Harapan (PKH) adalah suatu program yang dibuat oleh pemerintah dan diolah oleh Kementerian Sosial Republik Indonesia yang memiliki tujuan untuk mengatasi kemiskinan di daerah-daerah. Penelitian ini menggunakan tiga algoritma, C4.5 Decision Tree, Naïve Bayes dan K-Nearest Neighbour. Perhitungan dilakukan berdasarkan Accuracy, Precision, Recall dan kurva ROC. Tujuan penelitian ini adalah menentukan dan mengetahui perbandingan akurasi, presisi, nilai recall serta kurva ROC antara ketiga algoritma, sehingga ditemukan algoritma yang cocok dapat ditemukan dan digunakan dalam pengambilan keputusan bantuan Program Keluarga Harapan (PKH). Hasil yang di dapatkan yaitu algoritma K-NN yang memiliki akurasi terbaik dengan jumlah akurasi 98.34% dan nilai AUC yang dihasilkan sebesar 0.995, algoritma C4.5 memiliki akurasi 97.59%, AUC yang dihasilkan oleh kurva menghasilkan nilai sebesar 0.980 dan Naïve Bayes menghasilkan akurasi sebesar 90.65%, nilai AUC yang dihasilkan sebesar 0.965.

Kata kunci:

data mining; decision tree; naïve bayes; K-NN; program keluarga harapan

MERCU BUANA

ABSTRACT

Name : Monica Chelsea Ester Dondokambey
Student Number : 41516010136
Counsellor : Dr. Ida Nurhaida, MT
Title : Classification of C4.5, Naïve Bayes and K-Nearest Neighbour Algorithm for Receiver of Program Keluarga Harapan (PKH) DKI Jakarta Region

Poverty and inequality is one of the problems that occur in several countries. Indonesia is one of the countries in Southeast Asia that has these poverty problems. Program Keluarga Harapan (PKH) is a program created by the government and processed by the Ministry of Social of the Republic of Indonesia with a purpose to overcoming poverty in the regions. This research uses three algorithms, C4.5 Decision Tree, Naïve Bayes and K-Nearest Neighbor. Calculations are based on Accuracy, Precision, Recall and ROC curves. The purpose of this study is to determine the comparison of accuracy, precision, recall values and ROC curves between the three algorithms, so that a suitable algorithm is find and used in decision of the Program Keluarga Harapan (PKH). The results obtained are the K-NN algorithm that has the best accuracy, with a total accuracy of 98.34% and the resulting AUC value is 0.995, the C4.5 algorithm has an accuracy of 97.59%, the AUC generated by the curve produces a value of 0.980 and Naïve Bayes produces an accuracy of 90.65%, the resulting AUC value was 0.965.

Key words:

data mining; decision tree; naïve bayes; K-NN; program keluarga harapan

UNIVERSITAS
MERCU BUANA

KATA PENGANTAR

Puji syukur kita panjatkan kepada Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour Untuk Penerima Bantuan Program Keluarga Harapan (PKH) Wilayah DKI Jakarta” tepat pada waktunya. Tugas akhir ini disusun untuk memenuhi salah satu syarat memperoleh gelar sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Penulis menyadari bahwa tanpa bantuan dan bimbingan yang melibatkan banyak pihak, penelitian ini tidak akan terlaksana dengan baik. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Ibu Dr. Ida Nurhaida, MT, selaku dosen pembimbing tugas akhir yang telah meluangkan waktunya untuk memberikan bimbingan serta arahan dalam penyusunan tugas akhir ini hingga selesai.
2. Ibu Desi Ramayanti, S.Kom., MM, selaku Ka. Prodi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.
3. Bapak Diky Firdaus, S.Kom., MM, selaku Koordinator Tugas Akhir Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana tahun ajaran 2019/2020.
4. Orang tuaku (alm papi, mami, dan adik) yang senantiasa memberikan doa dan dukungan kepada penulis.
5. Teman-teman yang selalu menyemangati dan memberi motivasi kepada penulis selama pelaksanaan tugas akhir.
6. Semua pihak yang telah banyak membantu dalam penyusunan tugas akhir ini yang tidak dapat penulis sebutkan satu persatu.

Akhir kata, penulis berharap tugas akhir ini dapat bermanfaat bagi pembaca guna menambah pengetahuan dan wawasan.

Jakarta, 17 Februari 2020
Penulis

Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour Untuk Penerima Bantuan Program Keluarga Harapan (PKH) Wilayah DKI Jakarta

Classification of C4.5, Naïve Bayes and K-Nearest Neighbour Algorithms for Receiver of Program Keluarga Harapan (PKH) DKI Jakarta Region

Fildzah Gita Oktoviana¹⁾, Monica Chelsea Ester Dondokambey²⁾, Ida Nurhaida³⁾

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana
Jl. Raya Meruya Selatan, Kembangan, Jakarta, Indonesia 11650

Cara sitasi: F. G. Oktoviana, M. C. E. Dondokambey and I. Nurhaida, "Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour Untuk Penerima Bantuan Program Keluarga Harapan (PKH) Wilayah DKI Jakarta" *Jurnal Teknologi dan Sistem Komputer*, vol. x, no. x, 201x. doi: 10.14710/jtsiskom.x.x.xxxx.xx-xx, [Online].

Abstract - Poverty and inequality is one of the problems that occur in several countries. Indonesia is one of the countries in Southeast Asia that has these poverty problems. Program Keluarga Harapan (PKH) is a program created by the government and processed by the Ministry of Social of the Republic of Indonesia with a purpose to overcoming poverty in the regions. This research uses three algorithms, C4.5 Decision Tree, Naïve Bayes and K-Nearest Neighbor. Calculations are based on Accuracy, Precision, Recall and ROC curves. The purpose of this study is to determine the comparison of accuracy, precision, recall values and ROC curves between the three algorithms, so that a suitable algorithm is found and used in decision of the Program Keluarga Harapan (PKH). The results obtained are the K-NN algorithm that has the best accuracy, with a total accuracy of 98.34% and the resulting AUC value is 0.995, the C4.5 algorithm has an accuracy of 97.59%, the AUC generated by the curve produces a value of 0.980 and Naïve Bayes produces an accuracy of 90.65%, the resulting AUC value was 0.965.

Keywords - data mining; decision tree; naïve bayes; K-NN; program keluarga harapan

Abstrak - Kemiskinan dan ketimpangan merupakan salah satu permasalahan yang terjadi di beberapa negara. Indonesia merupakan salah satu negara di Asia Tenggara yang memiliki permasalahan kemiskinan tersebut. Program Keluarga Harapan (PKH) adalah suatu program yang dibuat oleh pemerintah dan diolah oleh Kementerian Sosial Republik Indonesia yang memiliki tujuan untuk mengatasi kemiskinan di daerah-daerah. Penelitian ini menggunakan tiga algoritma, C4.5 Decision Tree, Naïve Bayes dan K-Nearest Neighbour. Perhitungan dilakukan berdasarkan Accuracy, Precision, Recall dan kurva ROC. Tujuan penelitian ini adalah menentukan dan mengetahui perbandingan akurasi, presisi, nilai recall serta kurva ROC antara ketiga

algoritma, sehingga ditemukan algoritma yang cocok dapat ditemukan dan digunakan dalam pengambilan keputusan bantuan Program Keluarga Harapan (PKH). Hasil yang di dapatkan yaitu algoritma K-NN yang memiliki akurasi terbaik dengan jumlah akurasi 98.34% dan nilai AUC yang dihasilkan sebesar 0.995, algoritma C4.5 memiliki akurasi 97.59%, AUC yang dihasilkan oleh kurva menghasilkan nilai sebesar 0.980 dan Naïve Bayes menghasilkan akurasi sebesar 90.65%, nilai AUC yang dihasilkan sebesar 0.965.

Kata kunci - data mining; decision tree; naïve bayes; K-NN; program keluarga harapan

I. PENDAHULUAN

Kemiskinan dan ketimpangan merupakan salah satu permasalahan yang terjadi di beberapa negara. Indonesia merupakan salah satu negara di Asia Tenggara yang memiliki permasalahan kemiskinan tersebut [1]. Pada tahun 2017, persentase penduduk Indonesia yang hidup dalam kemiskinan tercatat sebesar 10.1% [2]. Untuk meminimalisir kemiskinan, maka pemerintah melakukan berbagai upaya dengan memberikan bantuan berupa uang tunai yang dinamakan Program Keluarga Harapan (PKH). Program Keluarga Harapan (PKH) adalah suatu program yang dibuat oleh pemerintah dan diolah oleh Kementerian Sosial Republik Indonesia yang memiliki tujuan untuk mengatasi kemiskinan di daerah-daerah.

Program bantuan tersebut telah diselenggarakan pemerintah sejak tahun 2007 [3]. Dengan bantuan tersebut di harapkan tingkat kemiskinan di Indonesia menurun. Program bantuan tersebut berjalan di bawah naungan Kementerian Sosial Republik Indonesia (KEMENSOSRI) dan dibantu oleh Dinas Sosial DKI Jakarta dalam pendataan anggota-anggota RTSM (Rumah Tangga Sangat Miskin). Meskipun penerima bantuan mendapatkan uang tunai, namun untuk mendapatkan bantuan ini dikenakan beberapa persyaratan yang digunakan untuk menyaring calon penerima bantuan PKH ini.

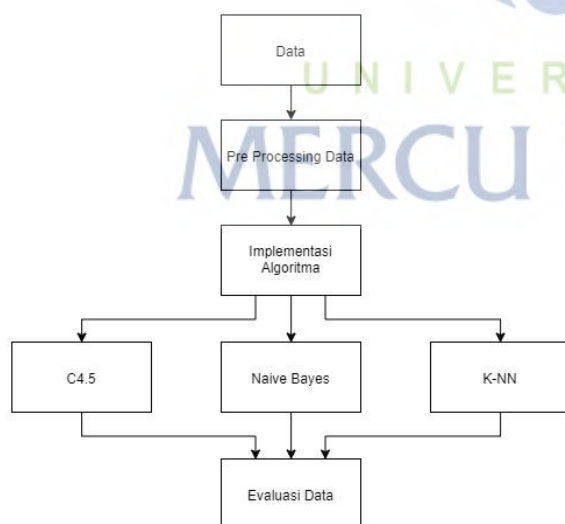
Dalam penentuan tersebut, ada beberapa anggota keluarga yang tidak mendapatkan, meskipun seharusnya

*) Penulis korespondensi (Ida Nurhaida)
Email: ida.nurhaida@mercubuana.ac.id

mendapatkan bantuan tersebut. Ada juga anggota keluarga yang termasuk mampu, namun malah mendapatkan bantuan PKH. Sehingga penyebaran bantuan tersebut tidak tepat sasaran [4]. Dalam permasalahan tersebut dilakukanlah penelitian mengenai data mining, dengan mengimplementasi ketiga algoritma terhadap permasalahan tersebut. Tugas utama data mining adalah menganalisis dan mengekstrak jumlah data yang besar dan sebelumnya belum pernah di teliti secara otomatis dan semi otomatis, sehingga menciptakan pola yang menarik dan kelompok dari kumpulan data [5]. Salah satu metode yang dapat diterapkan ialah klasifikasi yang digunakan untuk membantu dalam pengambilan keputusan [6]. Sasaran klasifikasi adalah menemukan fungsi klasifikasi atau model, yang dapat memetakan satu catatan dalam database ke kelas-kelas pra-asumsi [7]. Beberapa contoh metode klasifikasi ialah *C4.5 Decision Tree*, *Naïve Bayes*, *K-Nearest Neighbour (K-NN)*, *Logistic Regression* dan *Neural Network*. Pada penelitian ini, peneliti mencoba mengimplementasikan bantuan Program Keluarga Harapan ke dalam data mining menggunakan algoritma *C4.5*, *Naïve Bayes* dan *K-NN*. Setelah itu dilakukan perhitungan berdasarkan *Accuracy*, *Precision*, *Recall* dan *ROC Curve* yang di dapatkan. Sehingga mendapatkan kesimpulan algoritma mana yang terbaik dan tepat dalam kasus Program Keluarga Harapan tersebut.

II. METODE PENELITIAN

Tahapan yang dilakukan untuk menyelesaikan penelitian ditunjukkan oleh Gambar 1.



Gambar 1. Tahapan penelitian

Penelitian ini terdiri dari 5 tahapan, yaitu pengumpulan data, pre-processing data, implementasi algoritma dan evaluasi data.

1. Pengumpulan Data

Data yang didapatkan merupakan data kemiskinan yang terdapat di dalam Basis Data Terpadu. Data tersebut

di dapatkan dari Kementerian Nasional Republik Indonesia (KEMENSOSRI). Data yang diambil sekitar 2652 data dan memiliki 50 atribut total. Terdapat data sintesis atau data tambahan berupa atribut pendapatan yang di dapatkan dari Badan Pusat Statistik (BPS). Pada penelitian ini hanya diambil beberapa atribut yang mengacu kepada persyaratan yang di tetapkan untuk mendapatkan bantuan (Tabel 1).

Beberapa atribut yang dipilih tersebut berdasarkan kriteria dalam penentuan penerima bantuan PKH. Kriteria tersebut yaitu:

- 1) Adanya ibu hamil
- 2) Adanya ibu nifas
- 3) Balita
- 4) Anak usia dibawah 6 tahun
- 5) Anak SD/SMP/SMA berumur 7 -18
- 6) Penyandang disabilitas
- 7) Lanjut usia berumur 60 tahun keatas
- 8) Anak berusia 6 – 21 tahun yang belum menyelesaikan wajib belajar 12 tahun
- 9) Keluarga tidak mampu dan berpenghasilan rendah

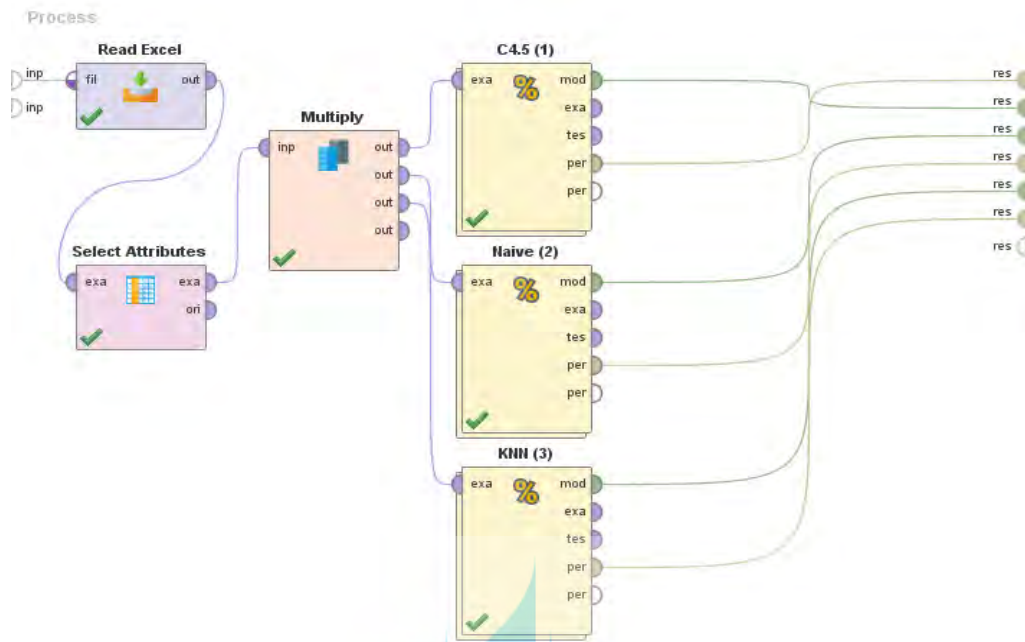
2. Pre-Processing Data

Pre-processing, terdapat dua tahap yang dilakukan. Tahap pertama dari *pre-processing* data adalah tahap cleaning atau pembersihan terhadap data yang terdapat missing values serta pemilihan beberapa atribut menggunakan operator *select attribute* yang dijadikan sebagai parameter pembanding dalam pengambilan keputusan (Tabel 1). Teknik Klasifikasi Algoritma dimulai dengan processing dan transformasi data agar data mentah yang digunakan untuk analisa adalah data dengan atribut yang lengkap [8].

Tabel 1. Daftar atribut-atribut yang digunakan

No	Atribut	Tipe Data
1	ada_ac	Binominal
2	ada_motor	Binominal
3	Umur	Integer
4	Jenis_disabilitas	Binominal
5	Partisipasi_sekolah	Polynomial
6	Pendidikan_tertinggi	Polynomial
7	Sta_Bekerja	Binominal
8	Pendapatan	Polynomial
9	Mendapat_PKH	Binominal

Data tambahan yang didapatkan dari Badan Pusat Statistik (BPS). Pada saat data di dapatkan, data tersebut masih berbentuk angka pendapatan dari beberapa jenis pekerjaan. Oleh karena itu data tersebut di



Gambar 2. Proses modelling

transformasikan lagi menjadi 4 kategori yang berbeda (Tabel 2). Setelah data di transformasikan lalu data sintesis tersebut digabungkan dengan atribut-atribut pada data utama (Tabel 1).

Tabel 2. Transformasi data pendapatan BPS

No	Pendapatan	Nilai Pre-Processing
1	>3.500.000	Sangat Tinggi
2	2.500.000 – 3.500.000	Tinggi
3	1.500.000 – 2.500.000	Sedang
4	<1.500.000	Rendah

3. Implementasi Algoritma

Tahapan selanjutnya adalah pengimplementasian algoritma yang telah melewati tahap pre-processing. Data di implementasikan ke dalam model algoritma C4.5, Naïve Bayes dan K-NN menggunakan 10-fold cross validation (Gambar 2).

3.1 Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma di dalam klasifikasi yang digunakan untuk membuat pohon keputusan (*Decision Tree*) [9]. *Decision tree* merupakan salah satu metode klasifikasi tradisional yang memiliki keuntungan memiliki struktur model yang simple, mudah untuk dipahami dan proses pengujiannya [10].

Secara umum, langkah-langkah algoritma C4.5 dalam membangun *decision tree* adalah sebagai berikut [11]:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk setiap nilai.
3. Bagi kasus dalam cabang.

4. Ulangi proses untuk setiap proses cabang sampai semua kasus pada cabang memiliki kelas yang sama.

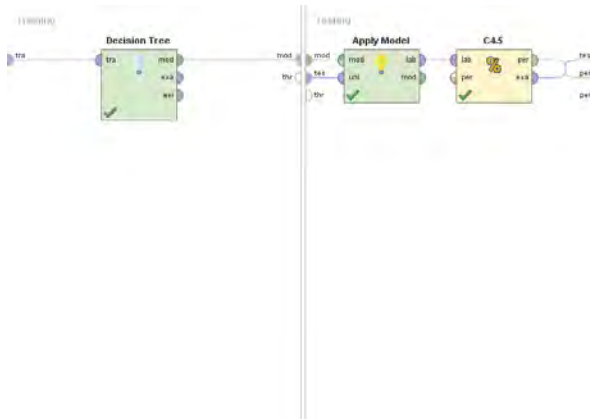
3.2 Naïve Bayes

Naïve Bayes adalah teknik prediksi berdasarkan probabilitas sederhana berdasarkan dalil *Bayes*. Penggunaan teorema *Bayes* pada metode *Naïve Bayes* adalah dengan menggabungkan probabilitas sebelumnya dan probabilitas bersyarat dalam formula yang dapat digunakan untuk menghitung probabilitas setiap kemungkinan klasifikasi [12]. Keuntungan dari *Naïve Bayes* adalah algoritma ini hanya memerlukan jumlah data uji yang sedikit untuk memperkirakan variasi dari variable yang dibutuhkan untuk klasifikasi [13].

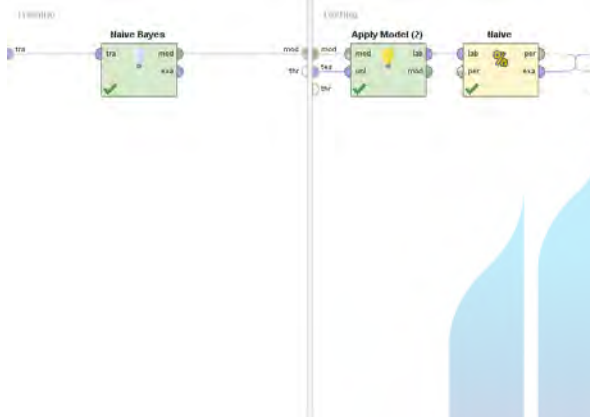
3.3 K-Nearest Neighbour

K-Nearest Neighbour adalah algoritma yang paling simple diantara algoritma lainnya [14]. *K-Nearest Neighbour (K-NN)* adalah sebuah metode supervised yang berarti membutuhkan data training untuk mengklasifikasikan objek yang jaraknya paling dekat. Prinsip kerja *K-Nearest Neighbour* adalah mencari jarak terdekat antara data yang akan di evaluasi dengan k tetangga (*neighbour*) dalam data pelatihan [15].

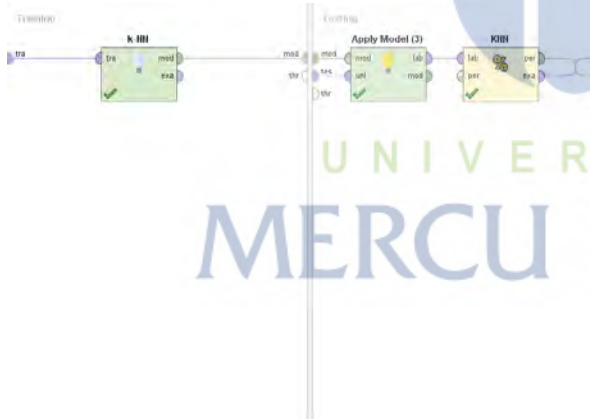
Cross Validation merupakan salah satu teknik untuk menilai/memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu [16]. Pada penelitian ini, masing-masing algoritma di terapkan ke dalam *cross validation* (Gambar 2),(Gambar 3),(Gambar 4).



Gambar 3. Cross validation algoritma C4.5



Gambar 4. Cross validation algoritma Naive Bayes



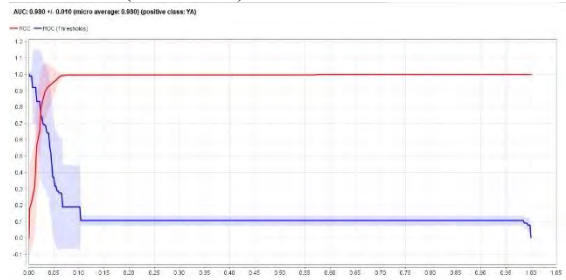
Gambar 5. Cross validation algoritma K-NN

Tahap yang dilakukan selanjutnya adalah evaluasi data menggunakan confusion matrix. Ketiga algoritma yang telah di uji tersebut lalu dibandingkan berdasarkan accuracy, precision, recall. Confusion Matrix adalah model yang akan membentuk matrix yang terdiri dari true positif atau tupel positif dan true negatif atau tupel negatif [17].

III. HASIL DAN PEMBAHASAN

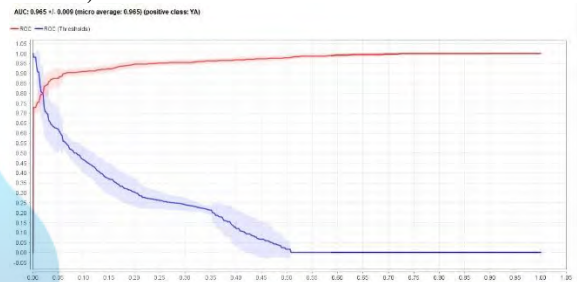
Algoritma C4.5 dalam klasifikasi penerimaan bantuan PKH menghasilkan accuracy sebesar 97.59%, presisi sebesar 96.32%, dan recall sebesar 98.88% (Tabel

5). AUC yang dihasilkan oleh kurva menghasilkan nilai sebesar 0.980 (Gambar 5).



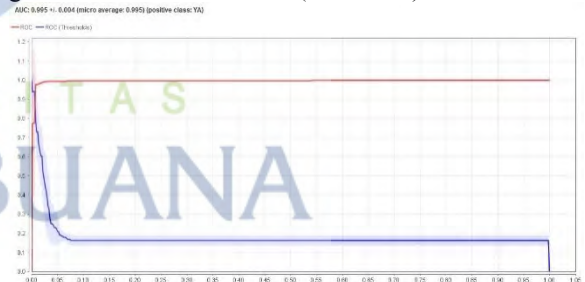
Gambar 5. AUC algoritma C4.5

Untuk algoritma *Naive Bayes* menghasilkan *accuracy* sebesar 90.65%, *presisi* 90.55% dan *recall* 90.21% (Tabel 5). Nilai AUC yang dihasilkan sebesar 0.965 (Gambar 6).



Gambar 6. AUC algoritma Naive Bayes

Algoritma terakhir yaitu *K-NN* yang menghasilkan akurasi terbesar di antar lainnya, yaitu 98.34%, *presisi* 98.20% dan *recall* sebesar 98.36% (Tabel 5). Nilai AUC yang dihasilkan sebesar 0.995 (Gambar 7).



Gambar 5. AUC algoritma K-NN

Pengujian dilakukan dalam algoritma *K-NN* sebanyak 6 kali dengan nilai *k* yang berbeda-beda, yaitu bernilai 1, 3, 5, 7, 9 dan 11. Bahwa nilai *k* terbaik adalah *k* = 5 (Tabel 3). Dalam penelitian menggunakan *k* bernilai 5.

Tabel 3. Perbandingan Nilai k pada algoritma KNN

NILAI K	NILAI ACCURACY (%)	NILAI PRECISION (%)	NILAI RECALL (%)
1	98,27%	98,03%	98,36%
3	98,34%	97,97%	98,60%
5	98,34%	98,20%	98,36%
7	98,23%	98,11%	98,20%
9	98,19%	98,19%	98,03%
11	98,19%	98,11%	98,10%

Tabel 4. Confusion matrix algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour

Algoritma Klasifikasi	TP	TN	FP	FN
C4.5	1314	1274	49	15
Naïve Bayes	1241	1163	122	126
K-NN	1340	1268	23	21

Tabel 5. Perbandingan kinerja algoritma C4.5, Naive Bayes dan K-NN

	Accuracy	Precision	Recall	AUC
C4.5	97.59%	96.32%	98.88%	0.985
K-NN	98.34%	98.20%	98.36%	0.995
Naïve Bayes	90.65%	90.55%	90.21%	0.965

Perbandingan kinerja ketiga algoritma yang dipakai menunjukkan akurasi diatas 90%, dengan akurasi tertinggi diraih oleh algoritma K-NN yaitu 98.34%. Namun algoritma K-NN memiliki selisih yang tidak terlalu jauh dari algoritma C4.5 yang bernilai 97.59. Naïve Bayes memiliki hasil sebesar 90.65% (Tabel 5).

Dari hasil yang terdapat di (Tabel 5) Algoritma K-NN merupakan algoritma yang memiliki hasil tertinggi. Hasil tersebut di dapatkan karena data yang digunakan saling berkaitan satu sama lain. Algoritma K-NN bersifat non-linear dan menghitung berdasarkan jarak terdekat antara

data yang akan di evaluasi dengan k tetangga (*neighbour*) dalam data pelatihan.

Algoritma C4.5 memiliki hasil yang cukup baik dalam klasifikasi ini. Atribut yang digunakan dalam penelitian memiliki keterkaitan antar satu sama lain. Sehingga *decision tree* yang didapatkan bersifat baik dan mudah untuk dimengerti. Meskipun terdapat beberapa atribut yang pada akhirnya tidak digunakan dalam pembentukan akhir *decision tree*.

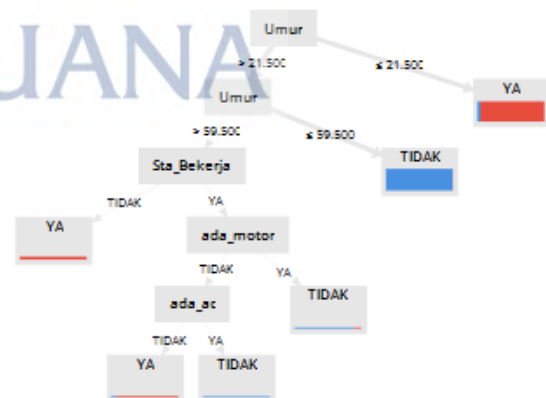
Naive Bayes mendapatkan hasil akurasi yang jauh dari kedua algoritma lainnya dikarenakan data yang digunakan dalam penelitian bersifat dependen. Naive Bayes akan bekerja lebih baik apabila data yang digunakan bersifat independen.

Berdasarkan penilaian AUC yang didapat dari masing-masing algoritma menunjukkan bahwa ketiga algoritma tersebut masuk ke dalam kategori Sangat Baik, yaitu ketiganya diatas nilai 0.90 dan hampir mendekati nilai 1.0. Kriteria penilaian nilai AUC di tampilkan pada (Tabel 6).

Tabel 6. Kriteria penilaian AUC

Nilai AUC	Interpretasi
0.90 – 1.00	Klasifikasi sangat baik
0.80 – 0.90	Klasifikasi baik
0.70 – 0.80	Klasifikasi sedang
0.60 – 0.70	Klasifikasi buruk
0.50 – 0.60	Klasifikasi gagal

Pohon keputusan yang dihasilkan oleh C4.5 dapat dilihat pada (Gambar 8).



Gambar 6. Pohon keputusan

Decision Tree yang dihasilkan oleh C4.5 dapat dilihat di (Gambar 8).

Dari pohon keputusan yang telah terbentuk pada (Gambar 8), di dapatkan beberapa aturan/rule model. Aturan-aturan tersebut adalah:

1. If Umur <21 Then = YA
2. If Umur >21 And Umur < 60 Then = Tidak

3. If Umur >21 And Umur >60, Sta_Bekerja = TIDAK, Then = YA
4. If Umur >21 And Umur >60, Sta_Bekerja = YA, ada_motor = YA, Then = TIDAK
5. If Umur >21 And Umur >60, Sta_Bekerja = YA, ada_motor = TIDAK, ada_ac = TIDAK, Then = YA
6. If Umur >21 And Umur >60, Sta_Bekerja = YA, ada_motor = TIDAK, ada_ac = YA, Then = TIDAK

Dari rule yang didapatkan, dapat disimpulkan:

1. Apabila calon penerima memiliki umur dibawah 21 tahun maka berhak mendapatkan bantuan.
2. Apabila calon penerima memiliki umur di antara umur 22 hingga 59 maka tidak berhak mendapatkan bantuan.
3. Apabila calon penerima memiliki umur diatas 21 tahun yaitu berusia 60 tahun keatas, serta tidak bekerja maka berhak mendapatkan bantuan.
4. Apabila calon penerima memiliki umur diatas 21 tahun yaitu berusia 60 tahun keatas, memiliki pekerjaan dan memiliki motor maka tidak berhak mendapatkan bantuan.
5. Apabila calon penerima memiliki umur diatas 21 tahun yaitu berusia 60 tahun keatas, memiliki pekerjaan, tidak memiliki motor, tidak memiliki ac maka berhak mendapatkan bantuan.

IV. KESIMPULAN

Klasifikasi penerimaan bantuan PKH menghasilkan kinerja terbaik menggunakan algoritma C4.5 dengan akurasi 97.89%. Hasil yang sama pun di dapatkan dari beberapa penelitian lain yang menyatakan bahwa C4.5 merupakan algoritma yang memiliki nilai akurasi yang paling baik. Akurasi yang di dapatkan tergantung dari implementasi dan data yang digunakan. Algoritma K-NN mencapai akurasi terbaik pada nilai k = 5 dengan akurasi sebesar 97.36%. Dapat disimpulkan bahwa nilai k pada K-NN mempengaruhi *accuracy*, yang di dapat pada K-NN. Algoritma terakhir yaitu *Naïve Bayes* dengan akurasi sebesar 90.95% (Tabel 5). Sementara itu parameter yang paling berpengaruh di dalam pohon keputusan merupakan parameter umur. Di karenakan parameter tersebut di jadikan root atau akar dari pohon keputusan (Gambar 8).

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kementerian Sosial Republik Indonesia

(KEMENSOSRI) atas partisipasinya untuk memberikan data kemiskinan yang digunakan dalam pelaksanaan penelitian. Penulis juga mengucapkan terima kasih kepada pihak-pihak yang telah membantu penulis dari awal pembuatan penelitian hingga akhir penelitian.

DAFTAR PUSTAKA

- [1] N. Janah, "Sistem Pendukung Keputusan Untuk Menentukan Penerima Bantuan Program Keluarga Harapan (Pkh) Bidang Pendidikan Menggunakan Metode Simple Additive Weighting (Saw) Study Kasus : (Di Desa Toyoresmi Kecamatan Ngasem Kabupaten Kediri)," *Univ. Nusan. PGRI Kediri*, pp. 1–10, 2016.
- [2] D. Wintana, Hikmatulloh, N. Ichsan, J. J. Purnama, and A. Rahmawati, "KLASIFIKASI PENENTUAN PENERIMA MANFAAT PROGRAM KELUARGA HARAPAN (PKH)," *Kumpul. J. Ilmu Komput.*, vol. 06, no. 03, pp. 254–264, 2019.
- [3] R. Hasanah, "Decision Support System Validation Recipient Program Keluarga Harapan (PKH) in Wonosari District using AHP- TOPSIS Method," *J. Ilm. Teknol. dan Inf.*, vol. 5, no. 2, pp. 111–121, 2016.
- [4] R. Aprilawati and P. Tarigan, "Implementasi Data Mining Untuk Menentukan Keluarga Yang Layak Mendapat Kartu Pkh (Program Keluarga Harapan) Dengan Metode K-Means Clustering," *J. Pelita Inform.*, vol. 18, no. 1, pp. 69–73, 2019.
- [5] S. S. Alaoui, Y. Farhaoui, and B. Aksasse, "Classification Algorithms in Data Mining : A Survey," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* ©, vol. 3, no. 1, pp. 349–355, 2018.
- [6] M. Mirqotussa'adah, M. A. Muslim, E. Sugiharti, B. Prasetyo, and S. Alimah, "Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mendiagnosa Diabetes," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 8, no. 2, p. 135, 2017.
- [7] Q. Dai, C. Zhang, and H. Wu, "Research of Decision Tree Classification Algorithm in Data Mining," *Int. J. Database Theory Appl.*, vol. 9, no. 5, pp. 1–8, 2016.
- [8] Y. A. Fiandra, S. Defit, and Y. Yuhandri, "Penerapan Algoritma C4.5 untuk Klasifikasi Data Rekam Medis berdasarkan International Classification Diseases (ICD-10)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 2, p. 82, 2017.
- [9] A. Khoerunnisa, B. Irawan, and M. R. Rumani, "Analisis dan implementasi perbandingan algoritma c.45 dengan naive bayes untuk prediksi penawaran produk," *E-Proceeding Eng.*, vol. 3, no. 3, pp. 5029–5035, 2016.
- [10] J. Zhao, Y. Liu, and M. Hu, "Optimisation algorithm for decision trees and the prediction

- of horizon displacement of landslides monitoring,” *J. Eng.*, vol. 2018, no. 16, pp. 1698–1703, 2018.
- [11] A. Muzakir and R. A. Wulandari, “Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree,” *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016.
- [12] D. R. Ente, S. Arifin, Andreza, and S. A. Thamrin, “Comparison of C4.5 algorithm with naive Bayesian method in classification of Diabetes Mellitus (A case study at Hasanuddin University hospital Makassar),” *J. Phys. Conf. Ser.*, vol. 1341, no. 9, 2019.
- [13] M. S. D. Dr. S. Vijayarani1, “Liver Disease Prediction using SVM and Naïve Bayes Algorithms,” *Int. J. Sci. Eng. Technol. Res.*, vol. 4, no. 4, pp. 816–820, 2015.
- [14] S. D. Jadhav and H. P. Channe, “Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques,” *Int. J. Sci. Res.*, vol. 5, no. 1, pp. 1842–1845, 2016.
- [15] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, “Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 427, 2018.
- [16] S. Dewi, “Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan,” *J. Techno Nusa Mandiri*, vol. 13, no. 1, pp. 60–66, 2016.
- [17] M. F. Arifin and D. Fitrihanah, “Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada,” *J. Telekomun. dan Komput.*, vol. 8, no. 2, pp. 87–102, 2018.



LAMPIRAN KORESPONDENSI

#13681 Ringkasan

Ringkasan | Tinjauan | Pengeditan

Naskah

Penulis	Fildzah Gita Oktoviana, Monica Chelsea Ester Dondokambey, Ida Nurhaida
Judul	Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour Untuk Penerima Bantuan Program Keluarga Harapan (PKH) Wilayah DKI Jakarta
File Asli	<input type="text" value="13681-28410-2-SM.docx"/> 2020-02-23
Singkatan File Tambahan	Tidak Ada <input type="button" value="Tambah file tambahan"/>
Naskah	Fildzah Gita Oktoviana 
Tanggal dikumpulkan	Pebruari 23, 2020 - 05:46
Bagian	Articles UNIVERSITAS
Editor	Tidak ada yang ditugaskan
Saran reviewer yang ditunjuk	Amir Mahmud Husein, Universitas Prima Indonesia

Status

Status	Menunggu Penugasan
Dimulai	2020-02-23
Terakhir Dimodifikasi	2020-02-23

[JTSISKOM] Pernyataan Naskah Inbox x



Editor-in-Chief Jurnal Teknologi dan Sistem Komputer <jtsiskom@ce.undip.ac.id>

5:48 PM (0 minutes ago)



to me ▾

Yth. Fildzah Gita Oktoviana,

Terima kasih atas pengiriman naskah Anda berjudul "Klasifikasi Algoritma C4.5, Naive Bayes dan K-Nearest Neighbour Untuk Penerima Bantuan Program Keluarga Harapan (PKH) Wilayah DKI Jakarta" ke Jurnal Teknologi dan Sistem Komputer. Naskah Anda akan melewati tahap review sesuai kebijakan jurnal kami. Review akan dilaksanakan sekitar 4-8 minggu setelah pengiriman. Lihat <http://jtsiskom.undip.ac.id/index.php/jtsiskom/about/editorialPolicies#peerReviewProcess>.

Dengan sistem manajemen jurnal daring ini Saudara dapat men-tracking perkembangan dari proses review naskah dengan login ke situs jurnal ini dengan menggunakan username Anda.

Manuscript URL: <https://jtsiskom.undip.ac.id/index.php/jtsiskom/author/submission/13681>

Username: fildzahgita

Jika diterbitkan, hak cipta dari artikel adalah milik JTSiskom dan Departemen Sistem Komputer Universitas Diponegoro sebagai penerbit jurnal. Anda diminta untuk mengisi dan menandatangani Formulir Perjanjian Pengalihan Hak Cipta (unduh di <https://goo.gl/Grs9qr>). Kirim formulir lengkap setelah artikel Anda dinyatakan diterima untuk diterbitkan.

Jika Saudara mempunyai pertanyaan tentang sistem manajemen jurnal ini, silakan kontak kami

Terima kasih atas kontribusinya di jurnal Jurnal Teknologi dan Sistem Komputer sebagai tempat publikasi Saudara

Editor-in-Chief Jurnal Teknologi dan Sistem Komputer

Jurnal Teknologi dan Sistem Komputer

CURRICULUM VITAE



Name : Monica Chelsea Ester Dondokambey
Place and date of birth : Tangerang, 10 Februari 1999
Sex : Perempuan
Nationality : Indonesia
Marital Status : Belum Menikah
ID Number : 3674035002990006
Pasport Number : -
Email : chelseadondokambey10@gmail.com
Religion : Protestan
Address : Jl. Komp. Villa Bintaro Indah Blok B2 No. 19, Kec. Jombang, Kec. Ciputat, 15414, Tangerang Selatan
Phone Number/Mobile : 0878-4206-9766

EDUCATION

Elementary School : SD Negeri Lengkong Wetan 02
Secondary School : SMP Negeri 14 Tangerang Selatan
Senior High School : SMA Negeri 11 Tangerang Selatan
University : Universitas Mercu Buana
Degree Awarded : S1
Faculty : Ilmu Komputer
Title of thesis : Klasifikasi Algoritma C4.5, Naïve Bayes dan K-Nearest Neighbour Untuk Penerimaan Bantuan Program Keluarga Harapan (PKH) Wilayah DKI Jakarta
GPA : 3.56

SKILLS

Language : Indonesia & Inggris **Score Toefl** : -

ORGANIZATION EXPERIENCE

No. of Years	Title	Name of Organization

WORK EXPERIENCE

No. of Years	Position	Employer
2018	Kuesioner Magang	Pusat Penelitian Universitas Mercu Buana
2018	Staff Magang	Pusat Pengabdian Pada Masyarakat Universitas Mercu Buana

ACTIVITIES

NATIONAL/INTERNATIONAL

No. of Years	Position	Employer

ACHIEVEMENT

No. of Years	Position	Employer

HOBBIES

Sports : Berenang, Bulutangkis, Tenis dan Voli
Various :

Universitas Mercu Buana

LEMBAR ASISTENSI



UNIVERSITAS
MERCU BUANA

KARTU ASISTENSI

NAMA : Monica Chebea E D SEM/THN AKAD : 7/
 NIM : 4151601036 JENIS BIMBINGAN : Tugas Akhir
 FAKULTAS : Ilmu Komputer DOSEN PEMBIMBING : Dr. Ida Murhaida, MT
 PROGRAM STUDI : Teknik Informatika
 JUDUL :

NO.	TGL	KETERANGAN	PARAF	NO.	TGL	KETERANGAN	PARAF
1.	5/19 /9	- Konfirmasi Data - Deadline Preprocessing	f	7.	8/20 /1	- Revisi Jurnal.	f
2.	5/19 /10	- Konfirmasi data - Perambahan atribut - Perambahan data	f	8.	18/20 /1	- Revisi Jurnal	f
3.	5/19 /11	- progress data - Penulisan jurnal.	f	9.	22/20 /1	Pendaftaran jaman Sidang	f
4.	27/19 /11	- Pengiriman jurnal untuk direvisi.	f	10.	31/20 /1	- Revisi Jurnal.	f
5.	3/20 /1	- Revisi Jurnal	f				
6.	22/19 /11	- Progress Jurnal	f				



KEMENTERIAN SOSIAL REPUBLIK INDONESIA
 SEKRETARIAT JENDERAL
 JALAN SALEMBA RAYA NOMOR 28 JAKARTA PUSAT 10430
 TELEPON (021) 3103591 LAMAN: <http://www.kemosos.go.id>

BERITA ACARA
 SERAH TERIMA DATA TERPADU PENANGANAN FAKIR MISKIN DAN ORANG
 TIDAK MAMPU BERDASARKAN SK MENTERI SOSIAL NOMOR 71/HUK/2018
NOMOR 546/SJ-PDT/BA/2019

Pada hari ini, Kamis, tanggal Delapan, bulan Agustus, tahun Dua Ribu Sembilan Belas, kami masing-masing:

1. Nama : Ujang Taofik Hidayat
 Jabatan : Kepala Bidang Diseminasi Data
 selanjutnya disebut PIHAK KESATU, dan
2. Nama : Monica Chelsea Ester Dondokambey
 NIK : 3674035002990006
 Pekerjaan : Mahasiswa
 Alamat : Jl. Komp. Villa Bintaro Indah Blok B2 No.19 RT 004 RW 012
 Jombang, Ciputat, Tangerang Selatan, Banten 15414
 No. Telp./HP : 87842069766

selanjutnya disebut PIHAK KEDUA;

PIHAK KESATU telah menyerahkan Data Terpadu Penanganan Fakir Miskin dan Orang Tidak Mampu berdasarkan SK Menteri Sosial Nomor 71/HUK/2018 kepada PIHAK KEDUA dan PIHAK KEDUA telah menerima dari PIHAK KESATU Data Terpadu Penanganan Fakir Miskin dan Orang Tidak Mampu berdasarkan SK Menteri Sosial Nomor 71/HUK/2018 untuk keperluan menyelesaikan tugas akhir dalam rangka memenuhi persyaratan Kurikulum Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Mercu Buana.

Dengan diterimanya data tersebut, PIHAK KEDUA menyatakan dan sepakat untuk menyetujui bahwa penggunaan data akan mengikuti syarat-syarat yang ditentukan oleh Kementerian Sosial, yaitu:

1. hanya menggunakan data untuk keperluan seperti tersebut dalam Berita Acara Serah Terima ini;
2. Tidak akan menyebarkan data tersebut kepada instansi/orang lain;
3. melaporkan secara tertulis dengan memberikan hasil pemanfaatan Data Terpadu Kesejahteraan Sosial kepada Menteri Sosial paling lambat 2 (dua) bulan setelah data digunakan.

Berita acara ini dibuat dengan sesungguhnya untuk dipergunakan sebagaimana mestinya.

Dibuat di Jakarta

PIHAK KEDUA

Monica Chelsea Ester Dondokambey

PIHAK PERTAMA

Ujang Taofik Hidayat