



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**PENERAPAN ALGORITMA C4.5 UNTUK
MENENTUKAN KELAYAKAN PENERIMA
BANTUAN PROGRAM KELUARGA HARAPAN**

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik Pada
Jurusan Teknik Informatika

Oleh :

AHMAD MUHAJIR HASIBUAN
11551102718



UIN SUSKA RIAU

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM
RIAU
PEKANBARU
2021**



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN

**PENERAPAN ALGORITMA C4.5 UNTUK MENENTUKAN
KELAYAKAN PENERIMA BANTUAN PROGRAM
KELUARGA HARAPAN**


TUGAS AKHIR

Oleh

AHMAD MUHAJIR HASIBUAN
11551102718

Telah diperiksa dan disetujui sebagai laporan tugas akhir
di Pekanbaru, pada tanggal 16 Agustus 2021

Pembimbing I,


Dr. Okfalisa, ST, M.Sc
NIP. 19771028 200312 2 004

ii

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

PENERAPAN ALGORITMA C4.5 UNTUK MENENTUKAN KELAYAKAN PENERIMA BANTUAN PROGRAM KELUARGA HARAPAN

TUGAS AKHIR

Oleh

AHMAD MUHAJIR HASIBUAN

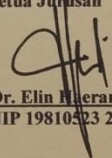
11551102718

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar sarjana Teknik Informatika
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 15 Juli 2021

Pekanbaru, 15 Juli 2021

Mengesahkan,

Ketua Jurusan


Dr. Elin Haerani, ST., M.Kom.
NIP 19810423 200710 2 003



DEWAN PENGUJI

Ketua : Iwan Iskandar, ST, MT
Sekretaris : Dr. Okfalisa, ST, M.Sc
Penguji I : Dr. Alwis Nazir, M.Kom
Penguji II : Elvia Budianita, ST, M.Cs



LEMBAR HAK KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. pada perpustakaan diperbolehkan untuk dicatat, tetapi pengutipan hanya dapat dilakukan dengan izin penulis dan wajib disertai dengan kebiasaan ilmiah untuk menyebutkan sumber-sumbernya.

Duplikasi atau penerbitan sebagian atau seluruh Tugas Akhir ini harus mendapatkan izin dari Dekan Fakultas Sains Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang ingin meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi data seperti nama, tanda peminjaman dan tanggal penelitian ini dipinjamkan.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LEMBAR PERNYATAAN

Dengan ini saya menyatakan bila dalam Tugas Akhir ini tidak mengandung karya yang pernah diajukan sebelumnya untuk memperoleh gelar sarjana di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacukan dalam naskah ini dan disebutkan di daftar pustaka.

Pekanbaru, 16 Agustus 2021

Yang membuat pernyataan,

AHMAD MUHAJIR HASIBUAN
11551102718

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Syukur, sembah dan sujud hanyalah kepada Allah SWT. Nikmat ilmu dan Entamu telah memberikanku kekuatan dalam menjalani hidup. Memperkenalkan aku dengan ilmu juga memperkenalkanku dengan cinta. Atas karunia serta kemudahan yang telah Engkau berikan akhirnya tugas akhir/skripsi yang sangat sederhana ini dapat terselasaikan dan berjalan dengan baik. Shalawat dan salam selalu terlimpahkan pada baginda Rasulullah Muhammad SAW. Kupersembahkan karya sederhana ini kepada orang yang sangat kusayangi dan kukasihi.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

PENERAPAN ALGORITMA C4.5 UNTUK MENENTUKAN KELAYAKAN PENERIMA BANTUAN PROGRAM KELUARGA HARAPAN

AHMAD MUHAJIR HASIBUAN
11551102718

Jurusan Teknik Informatika
Fakultas Sains Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Kemiskinan merupakan permasalahan multi-dimensi yang berkaitan dengan ketidakmampuan seseorang secara ekonomi, sosial budaya, politik dan bermasyarakat sosial. Untuk menanggulangi kemiskinan, pemerintah melakukan berbagai upaya dengan mengadakan program bantuan seperti program keluarga harapan (PKH). Program keluarga harapan (PKH) merupakan program kesejahteraan sosial bersyarat pemerintah yang diperuntukkan kepada masyarakat miskin yang telah terdaftar dalam basis data terpadu yang memenuhi syarat dan kriteria. Pada penelitian ini dilakukan klasifikasi data mining menggunakan algoritma C4.5 untuk menentukan kelayakan penerima bantuan program keluarga harapan (PKH). Langkah-langkah dalam penelitian ini adalah pengumpulan data yang didapat dari dinas sosial kabupaten rokan hulu. selanjutnya *selection data* yaitu dengan kriteria jumlah balita, ibu hamil, anak sd, anak smp, anak sma, pekerjaan, lanjut usia, disabilitas, status tempat tinggal, jenis dinding terluas, terima beras miskin, fasilitas jamban, tempat pembuangan akhir dengan kelas layak dan tidak layak. Setelah *selection data* maka dilakukan *pre-processing data* sebanyak 2244 *record* data yang mencakup pengecekan data kosong, data ganda dan kesalahan ketik. selanjutnya *transformastion* untuk memudahkan pengolahan data. Selanjutnya proses data mining dengan algoritma C4.5 hingga terbentuknya pohon keputusan. Pengujian menggunakan *Confusion Matrix* menghasilkan akurasi sebesar 86,37% dengan perbandingan 90% data latih dan 10% data uji. Berdasarkan hasil pengujian tersebut algoritma C4.5 termasuk pada range *Good Clasification*.

Kata kunci : *Algoritma C4.5, Kemiskinan, Klasifikasi, Pohon Keputusan, Program Keluarga Harapan (PKH).*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

APPLICATION OF THE C4.5 ALGORITHM TO DETERMINE THE ELIGIBILITY OF RECIPIENTS OF THE HOPE FAMILY PROGRAM HELP

AHMAD MUHAJIR HASIBUAN
11551102718

Informatics Engineering Departement
Faculty of Science and Technology
Islamic State University Sultan Syarif Kasim Riau

ABSTRACT

Poverty is a multi-dimensional problem related to the inability of a person economically, socio-culturally, politically and socially. To reduce poverty, the government has made various efforts by providing assistance programs such as the Family Hope Program (PKH). The Family Hope Program (PKH) is a conditional government social welfare program intended for the poor who have been registered in an integrated database that meets the requirements and criteria. In this study, data mining classification was carried out using the C4.5 algorithm to determine the eligibility of recipients of the Family Hope Program (PKH) assistance. The steps in this study were collecting data obtained from the social service of the Rokan Hulu district. Then the selection data is based on the criteria for the number of children under five, pregnant women, elementary school children, junior high school children, high school children, employment, elderly, disability, status of residence, widest type of wall, receiving poor rice, latrine facilities, final disposal site with decent class and not feasible. After data selection, 2244 data records were pre-processed which included checking blank data, duplicate data and typos. further transformation to facilitate data processing. Furthermore, the data mining process with the C4.5 algorithm until the formation of a decision tree. Testing using the Confusion Matrix produces an accuracy of 86.37% with a comparison of 90% of training data and 10% of test data. Based on the test results, the C4.5 algorithm is included in the Good Classification range.

Keywords: *C4.5 Algorithm, Poverty, Classification, Decision Tree, Family Hope Program (PKH).*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Assalamu'alaikum wa rahmatullahi wa barakaatuh

Segala puji dan syukur alhamdulillah atas kehadiran Allah Subhanaullahu Wa Ta'ala, yang telah melimpahkan segala nikmat, rahmat, hidayah dan kemudahannya kepada penulis sehingga dapat menyelesaikan tugas akhir yang berjudul “Penerapan Algoritma c4.5 untuk menentukan kelayakan penerima bantuan program keluarga harapan” ini dengan tepat waktu. Shalawat dan salam kepada Rasulullah Shalallahu ‘Alaihi Wa Sallam, yang telah membimbing ke jalan yang lurus, sehingga kita dapat merasakan ilmu-ilmu yang memudahkan aktifitas dan ibadah kita sehari-hari.

Tugas akhir ini ini dibuat sebagai salah satu syarat untuk mendapatkan gelar dari Sarjana Teknik jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau. Dalam proses menyelesaikan skripsi ini, penulis telah mendapatkan banyak sekali bimbingan, bantuan, dukungan serta motivasi baik secara langsung maupun tidak langsung. Untuk itu, pada kesempatan ini penulis ingin Ucapan terima kasih kepada:

Bapak Prof. Dr. Khairunnas, M.Ag selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau

Bapak Dr. Hartono, M.Pd selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.

DR. Elin Haerani, S.T, M.Kom sebagai Ketua Jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau.

Ibu Dr. Okfalisa, ST, M.Sc, selaku Pembimbing . Terimakasih untuk kesabaran, waktu, motivasi dan arahan yang telah diberikan sehingga tugas akhir ini dapat terselesaikan.

Bapak Dr. Alwis Nazir, M.Kom selaku Penguji I Tugas Akhir yang telah memberikan masukan dan arahan kepada penulis dalam penyusunan tugas akhir ini.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Ibu Elvia Budianita, ST, M.Cs. selaku penguji dua penulis yang telah memberikan kritik serta saran sebagai ilmu yang bermanfaat bagi penulis.

Seluruh Bapak dan Ibu yang mengajar di jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau

Untuk Ayahku Parmohonan Hasibuan dan Ibuku Damroh Terimakasih untuk perjuangannya sangat besar dan berat, agar aku dapat hidup bahagia, yang dimana gelar dan perjuangan keras ku ini aku persembahkan untuknya.

Untuk Abangku dan Kakakku dan Keluarga besarku, semua yang aku lakukan untuk kalian dan akan aku usahakan.

10. Teman-temanku Khairul Azmi, Ari Ismanto, Ary Bagus, Yusral Hadi, Ridho Ramidianto, Ridho Darma, Firman Syahrial, Fikry Sahputra dan Semua anggota keluarga besar TIF E 2015 dalam berjuang bersama.

11. Semua pihak yang terlibat langsung maupun tidak langsung dalam pelaksanaan tugas akhir ini yang tidak dapat disebutkan satu persatu.

Diharapkan tugas akhir ini bermanfaat bagi penulis khususnya maupun pembaca pada umumnya. Penulis sadar masih banyak sekali kekeliruan dan kekurangan oleh karena itu penulis berharap bisa mendapatkan masukan dari pembaca atas isi tugas akhir ini. Selamat membaca tugas akhir ini saya ucapkan terima kasih.

Pekanbaru, 16 Agustus 2021

Penulis

UIN SUSKA RIAU



DAFTAR ISI

BAB I	PENDAHULUAN.....	I-1
1.1	LATAR BELAKANG.....	I-1
1.2	RUMUSAN MASALAH	I-4
1.3	BATASAN MASALAH	I-4
1.4	TUJUAN PENELITIAN	I-4
1.5	SISTEMATIKA PENULISAN	I-5
BAB II	LANDASAN TEORI.....	II-6
2.1	KEMISKINAN.....	II-6
2.2	PROGRAM KELUARGA HARAPAN.....	II-7
2.3	DATA MINING	II-9
2.4	ALGORITMA C4.5	II-11
2.4.1	Pembuatan pohon keputusan.....	II-12
2.4.2	Kelebihan dan Kekurangan algoritma C4.5.....	II-15
2.5	<i>CONFUSION MATRIX</i>	II-16
2.6	PENELITIAN TERKAIT	II-17
BAB III	METODOLOGI PENELITIAN	III-21
3.1	Identifikasi Masalah	III-21
3.2	Studi Pustaka	III-22
3.3	Perumusan Masalah.....	III-22

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.4	Pengumpulan Data	III-22
3.5	Analisa.....	III-23
3.5.1	Analisa Kebutuhan Data	III-23
3.5.2	Analisa metode.....	III-24
3.6	Implementasi	III-24
3.7	Pengujian	III-25
3.8	Kesimpulan dan Saran.....	III-25
BAB IV ANALISA DAN PERANCANGAN.....		IV-26
4.1	Analisa Kebutuhan Data.....	IV-26
4.1.1	<i>Data Selection</i>	IV-26
4.1.2	<i>Pre-processing</i>	IV-30
4.1.3	Tranformation	IV-31
4.1.4	Data mining.....	IV-32
BAB V IMPLEMENTASI DAN PENGUJIAN		V-50
5.1	Implementasi	V-50
5.1.1	Lingkungan Implementasi	V-50
5.1.2	Import <i>Library</i> dan <i>Tools</i>	V-50
5.1.3	<i>Import Data</i>	V-51
5.1.4	<i>Transformation</i>	V-51
5.1.5	Implementasi Tahapan <i>Preprocessing</i>	V-52



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5.1.6	Implemtasi Klasifikasi	V-52
5.2	Pengujian	V-54
5.2.1	Pengujian dengan perbandingan 90%:10%	V-54
5.2.2	Pengujian dengan perbandingan 80%:20%	V-55
5.2.3	Pengujian dengan perbandingan 70%:30%	V-55
BAB VI PENUTUP		VI-56
6.1	Kesimpulan.....	VI-56
6.2	Saran.....	VI-56
Daftar pustaka		58

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB I PENDAHULUAN

1.1 LATAR BELAKANG

Masalah kesejahteraan sosial di Indonesia yang harus dikaji terus-menerus secara relevan adalah kemiskinan. Kemiskinan menjadi masalah utama yang harus segera ditanggulangi sebagaimana menjadi tujuan nasional yang tercantum dalam undang-undang dasar tahun 1945. Kemiskinan memang masalah yang selalu dihadapi manusia. Masalah kemiskinan memang sama tuanya dengan manusia itu sendiri dan implikasi permasalahannya dapat melibatkan berbagai segi kehidupan manusia. Kemiskinan merupakan permasalahan multi-dimensi yang berkaitan dengan ketidakmampuan seseorang, baik itu secara ekonomi, sosial budaya, politik dan bermasyarakat sosial (Nurwati, 2008). Dengan kata lain bahwa kemiskinan itu merupakan masalah sosial yang sifatnya mendunia dan sudah menjadi perhatian dunia meskipun dampak dari kemiskinan tersebut berbeda-beda. Berdasarkan data yang dirilis Badan Pusat Statistik pada tahun 2019 tercatat bahwa angka kemiskinan di Indonesia sebesar 9,22% dengan jumlah penduduk 24,74 juta orang pada september 2019, angka tersebut turun 0,36 juta orang terhadap bulan maret 2018 dan turun menjadi 0,88 juta orang pada bulan september 2018. Sedangkan jumlah penduduk miskin di kabupaten rokan hulu sebesar 72,21 ribu jiwa dengan persentase 10.53% penduduk kabupaten Rokan Hulu tercatat sebagai penduduk miskin ([BPS] Badan Pusat Statistik, 2019).

Banyaknya jumlah penduduk miskin, pemerintah berupaya menanggulangi masalah kemiskinan dengan berbagai program bantuan seperti: Bantuan langsung tunai (BLT), beras miskin (raskin), sembako dan program keluarga harapan (PKH). Program keluarga harapan (PKH) merupakan program kesejahteraan sosial bersyarat pemerintah yang diperuntukkan kepada masyarakat miskin yang telah terdaftar dalam basis data terpadu program penanganan fakir miskin yang memenuhi syarat dan kriteria. Kriteria peserta penerima manfaat (KPM) PKH

meliputi komponen pendidikan, kesehatan dan kesejahteraan sosial dengan kriteria yang telah ditetapkan yakni terdapat anak sekolah tingkat SD-SMA, ibu hamil, ibu menyusui, lansia dan disabilitas berat. Sebagai sebuah program bantuan sosial bersyarat, PKH membuka akses bagi keluarga miskin agar mampu memanfaatkan berbagai fasilitas layanan pendidikan, sosial dan fasilitas layanan kesehatan yang tersedia disekitar mereka guna untuk mempertahankan taraf kesejahteraan sosial sebagaimana sesuai dengan amanat konstitusi dan nawacita presiden RI.

Sebelum KPM program keluarga harapan mendapatkan bantuan, KPM terlebih dahulu melakukan tahap seleksi berupa verifikasi dan validasi data yang dilakukan oleh pendamping atau petugas yang mendata keluarga penerima manfaat program keluarga harapan tersebut. Pada tahap verifikasi dan validasi data yang dilakukan oleh petugas masih secara manual dengan mendatangi rumah keluarga penerima manfaat program keluarga harapan sehingga memakan waktu yang lama dalam menentukan keluarga yang layak mendapatkan bantuan program keluarga harapan. Maka dibutuhkan sebuah sistem simulasi yang dapat melakukan klasifikasi keluarga penerima manfaat bantuan program keluarga harapan dengan cepat. Dengan bantuan data mining sehingga memudahkan pihak instansi dalam menentukan keluarga yang masih layak atau tidak layak mendapatkan bantuan program keluarga harapan.

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi bermanfaat dan pengetahuan yang terakut dari berbagai database besar (Turban, 2005). Ada beberapa penelitian data mining yang digunakan untuk menganalisa kelayakan penerima bantuan program keluarga harapan Penelitian terkait program keluarga harapan yaitu penelitian tentang sistem pendukung keputusan kelayakan penerima program keluarga harapan (PKH) menggunakan algoritma *analytic network process*. Berdasarkan confusion matrix penelitian ini menghasilkan tingkat presisi

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

sistem pendukung keputusan sebesar 39,64%, *recall* sebesar 39,29% dan nilai *accuracy* sebesar 69,12%. Sedangkan pengujian fungsionalitas sistem menggunakan blackbox testing menghasilkan 93,34% sistem berjalan dengan baik (Anjarwati & Farahdibah, 2017).

Algoritma C4.5 merupakan salah satu dari jenis pohon keputusan yang sangat populer. Pohon keputusan sangat berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Algoritma C4.5 dapat menangani data numerik dan diskret dengan menggunakan rasio perolehan (*gain rasio*) Sebelum mendapatkan nilai rasio perolehan, perlu dilakukan perhitungan nilai informasi dalam satuan dari suatu kumpulan objek, yaitu dengan menggunakan konsep *entropy*. Dengan menggunakan pohon keputusan, peneliti tidak perlu melakukan estimasi pada distribusi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut. Karena metode ini menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan.

Berdasarkan penelitian sebelumnya terkait klasifikasi data mining dengan metode algoritma C4.5 yaitu Laily hermawanti (2012) yang berjudul “Penerapan algoritma klasifikasi C4.5 untuk diagnosis penyakit kanker payudara” menghasilkan nilai akurasi 94,56% dan nilai *Area Under Curve* (AUC) sebesar 0,941 sehingga menghasilkan hasil yang akurat. Penelitian lain dilakukan oleh Siti Masripah (2015) berjudul “evaluasi penentuan kelayakan pemberian kredit koperasi syariah menggunakan algoritma klasifikasi C4.5” menghasilkan nilai akurasi sebesar 88% dan nilai AUC sebesar 0,898 dengan tingkat diagnosa *good classification*. Kemudian penelitian selanjutnya dilakukan oleh Evicienna (2013) berjudul “algoritma C4.5 untuk prediksi hasil pemilihan legislatif DPRD DKI Jakarta” menghasilkan nilai akurasi yang tinggi yaitu sebesar 97.84% dan nilai AUC sebesar 0.970 dengan tingkat diagnosa *Excellent clasification*. Oleh karena itu, dapat disimpulkan algoritma C4.5 cocok diimplementasikan pada penelitian menentukan kelayakan menerima bantuan program keluarga harapan (PKH).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang yang telah diuraikan diatas maka dapat diambil suatu rumusan masalah dalam penelitian ini, yaitu: “penerapan algoritma C4.5 untuk menentukan kelayakan penerima bantuan program keluarga harapan (PKH) di kecamatan Rambah kabupaten Rokan Hulu”.

1.3 BATASAN MASALAH

Agar pembahasan penelitian ini tetap fokus, terstruktur dan terarah maka batasan penelitian tugas akhir ini adalah:

1. Pada algoritma C4.5 klasifikasi ditentukan oleh 2 kelas yaitu kelas layak dan tidak layak menerima bantuan program keluarga harapan
2. Data yang digunakan yaitu sebanyak 2244 data penerima program keluarga harapan kecamatan rambah kabupaten rokan hulu tahun 2017-2019.
3. Atribut yang digunakan pada penelitian ini yaitu jumlah balita, ibu hamil, anak sd, anak smp, anak sma, pekerjaan, lanjut usia, disabilitas, status tempat tinggal, jenis dinding terluas, terima beras miskin, fasilitas jamban, tempat pembuangan akhir.
4. Tools yang digunakan adalah microsoft excel dan bahasa pemograman phy-ton.

1.4 TUJUAN PENELITIAN

Adapun tujuan penelitian yang ingin dicapai pada penulisan tugas akhir ini adalah:

1. Menerapkan algoritma C4.5 untuk mengklasifikasikan kelayakan penerima bantuan program keluarga harapan berdasarkan kriteria-kriteria yang telah ditentukan.
2. Membangun sistem klasifikasi sebagai pendukung dalam menentukan kelayakan peserta penerima bantuan program keluarga harapan



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

SISTEMATIKA PENULISAN

Sistematika penulisan pada laporan tugas akhir ini terdiri dari beberapa bagian yang diuraikan dalam bentuk bab-bab sebagai berikut :

BAB I PENDAHULUAN

Bab ini menjelaskan tentang uraian latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan yang akan dibuat dalam tugas akhir ini.

BAB II LANDASAN TEORI

Bab ini akan menjelaskan tentang teori-teori umum dan khusus yang berhubungan dengan algoritma C4.5, data mining, klasifikasi, program keluarga harapan, *K-fold cross validation*, confusion matriks dan *rapid miner*. Teori yang didapatkan berasal dari buku dan jurnal tentang penelitian sejenis yang diakses melalui *internet*.

BAB III METODOLOGI PENELITIAN

Bab ini membahas langkah-langkah yang dilaksanakan dalam proses penelitian tugas akhir ini, yaitu mulai dari identifikasi masalah, perumusan masalah, studi pustaka, implementasi, pengujian, kesimpulan dan saran.

BAB IV ANALISA DAN PERANCANGAN

Bab ini berisikan tentang pembahasan mengenai analisa terhadap alur algoritma C4.5.

BAB V IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi implementasi sistem dan pengujian sistem. Kemudian kesimpulan terhadap sistem dari hasil pengujian yang dilakukan.

BAB VI PENUTUP

Dalam bab ini akan dijelaskan mengenai kesimpulan dari hasil penelitian yang telah dilakukan dan saran terhadap penelitian kedepannya.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II LANDASAN TEORI

Penulisan penelitian ini menggunakan berbagai referensi yang bersumber dari buku, jurnal, *E-book* dan *internet* untuk memahami tentang data mining, algoritma klasifikasi terkait penelitian serta referensi tentang bantuan program keluarga harapan (PKH).

2.1 KEMISKINAN

Kemiskinan merupakan permasalahan multi-dimensi yang berkaitan dengan ketidakmampuan seseorang, baik itu secara ekonomi, sosial budaya, politik dan bermasyarakat sosial (Nurwati, 2008). Kemiskinan adalah masalah yang kompleks, yang penanggulangannya tidak mudah. Terdapat bentuk-bentuk kemiskinan serta berbagai macam faktor kemiskinan di Indonesia. Kemiskinan tidak hanya disebabkan atau dipengaruhi oleh terbatasnya pendapatan, tetapi juga dipengaruhi tingkat pendidikan dan kesehatan. Sehingga peran pemerintah tidak hanya berusaha bagaimana agar masyarakat dapat meningkatkan pendapatannya, tetapi juga bagaimana dapat meringankan beban mereka khususnya bidang pendidikan dan kesehatan.

Ada dua ukuran kemiskinan yang umum digunakan saat ini yaitu kemiskinan absolut dan kemiskinan relatif (Kuncoro, 2006). Kemiskinan absolut berkaitan dengan perkiraan tingkat pendapatan dan kebutuhan dasar minimal untuk hidup layak. Tingkat pendapatan merupakan batas antara kondisi miskin atau tidak yang sering disebut dengan garis kemiskinan. Masyarakat yang memiliki pendapatan dibawah garis kemiskinan masuk dalam kelompok masyarakat miskin. Sedangkan kemiskinan relatif menggambarkan orang yang telah mencapai pendapatan sudah mencapai kebutuhan dasar minimum. Tetapi jika masih jauh dari pemenuhan kebutuhan hidup sederhana dalam lingkungannya maka orang tersebut masih dianggap miskin.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.2 PROGRAM KELUARGA HARAPAN

Program keluarga harapan (PKH) merupakan program pemberian bantuan bantuan sosial bersyarat yang diberikan kepada keluarga miskin yang ditetapkan sebagai keluarga penerima manfaat (KPM) (kemos, 2018). Dalam rangka meningkatkan efektivitas dan efisiensi penyaluran bantuan sosial serta mewujudkan prinsip 4T (Tepat Sasaran, Tepat Waktu, Tepat Jumlah, dan Tepat Administrasi) dan mendorong keuangan inklusif, Presiden Republik Indonesia (RI) memberikan arahan agar bantuan sosial dan subsidi disalurkan secara non tunai (Ratas tentang Keuangan Inklusif tanggal 26 April 2016). Melalui penyaluran bantuan sosial non tunai dengan menggunakan sistem perbankan, diharapkan dapat meningkatkan transparansi dan akuntabilitas program penyaluran bantuan sosial sehingga mudah dikontrol, dipantau dan mengurangi penyimpangan (Sosial, 2019). Tujuan pemberian bantuan program keluarga harapan adalah sebagai berikut:

1. Menciptakan perubahan perilaku dan kemandirian keluarga penerima manfaat dalam mengakses layanan kesehatan dan pendidikan serta kesejahteraan sosial yang akan meningkatkan taraf hidup keluarga penerima manfaat, melalui akses layanan pendidikan, kesehatan, dan kesejahteraan sosial.
2. Mengurangi beban pengeluaran dan meningkatkan pendapatan keluarga miskin dan rentan sehingga akan mengurangi kemiskinan dan kesenjangan.
3. Mengurangi kemiskinan dan mengenalkan produk dan jasa keangan formal kepada keluarga penerima manfaat.

Program keluarga harapan (PKH) diperuntukkan bagi keluarga miskin dan rentan terhadap resiko sosial yang telah terdaftar dalam basis data terpadu yang telah memenuhi kriteria dan komponen kesehatan, pendidikan, dan kesejahteraan sosial yang tinggal di pemukiman terpencil, wilayah pesisir/pulau kecil dan perbatasan. Penjelasan kriteria dan komponen penerima bantuan program keluarga harapan dapat dilihat pada tabel 2.1 dibawah ini.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Table 2.1 komponen dan kriteria penerima program keluarga harapan(PKH) (kemosos, 2018).

No	Komponen	Kriteria
	Pendidikan	<ul style="list-style-type: none"> - Anak SD atau sederajat - Anak SMP atau sederajat - Anak SMA atau sederajat - Anak usia 6-21 tahun yang belum menyelesaikan wajib belajar
	Kesehatan	<ul style="list-style-type: none"> - Ibu hamil - Ibu menyusui
3	Kesejahteraan sosial	<ul style="list-style-type: none"> - Lanjut usia ≥ 60 tahun - Penyandang disabilitas berat

Bantuan yang didapatkan penerima bantuan program keluarga harapan berbeda-beda tergantung pada komponen keluarga yang terdaftar. Berikut indeks yang diberikan pemerintah kepada masyarakat penerima bantuan program keluarga harapan dapat dilihat pada tabel 2.2 dibawah ini.

Table 2.2 indeks dan komponen bantuan PKH (kemosos, 2018)

No	Komponen bantuan	Indeks bantuan
1	Bantuan Tetap Reguler	Rp.550.000
2	Bantuan Tetap PKH Akses	Rp.1000.000
3	Bantuan Ibu Hamil	Rp.2.400.000
4	Bantuan Anak Usia Dini	Rp.2.400.000
5	Bantuan Anak SD	Rp.900.000



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

6	Bantuan Anak SMP	Rp.1.500.000
7	Bantuan Anak SMA	Rp.2.000.000
8	Bantuan Disabilitas Berat	Rp.2.400.000
9	Bantuan Lanjut Usia	Rp.2.400.000

Berdasarkan komponen dan kriteria penerima bantuan program keluarga harapan, keluarga yang telah terdaftar akan menjadi penerima bantuan apabila memenuhi salah satu komponen dan kriteria penerima bantuan yaitu komponen pendidikan, kesehatan dan kesejahteraan sosial. Kelayakan penerima program keluarga harapan juga di dukung oleh keterangan keadaan pokok rumah tangga yang telah di verifikasi dan di validasi oleh dinas sosial kriteria rumah tangga tersebut adalah pekerjaan, status tempat tinggal, jenis dinding terluas, terima beras miskin, fasilitas jamban, tempat pembuangan tinja akhir.

2.3 DATA MINING

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban, 2005). Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. (Larose, 2005). Sehingga data mining itu dapat diartikan sebagai proses penambangan data yang menghasilkan sebuah output (keluaran) berupa pengetahuan. Data mining dapat dibagi menjadi beberapa fase atau proses sesuai tugasnya masing-masing, yaitu (Dicky Nofriansyah, Gunadi Widi Nurcahyo, 2015):

Seleksi data (*Selection*)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Seleksi data merupakan pemilihan sekumpulan data operasional yang perlu dilakukan sebelum tahap penggalian informasi dalam data mining dimulai, disimpan dalam suatu berkas, terpisah dari basis data operasional.

Pemilihan data (Preprocessing/cleaning)

Proses pembersihan data mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Proses enrichment juga dilakukan yaitu proses memperkaya data yang sudah ada dengan data atau informasi yang relevan dan diperlukan untuk data mining.

3. Transformasi (*Transformation*)

Pada fase ini yang dilakukan adalah mentransformasi bentuk data yang belum memiliki entitas yang jelas kedalam bentuk data yang valid atau siap untuk melakukan data mining.

4. Data mining

Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode pencarian pengetahuan.

5. Interpretasi/evaluasi (*Interpratation/evaluation*)

Pada fase terakhir ini dilakukan adalah proses pembentukan keluaran yang sudah dimengerti bersumber pada proses data mining pola informasi.

Data mining terbagi menjadi beberapa kelompok dalam proses pemecahan masalah dan pencarian pengetahuan baru sesuai dengan tugasnya masing-masing, yaitu (Mardi, 2015):

Clasification

Clasification merupakan metode Data Mining yang bertugas untuk mengelompokkan pada setiap keadaan dimana tiap keadaan tersebut memiliki sekelompok atribut yang memiliki class. Metode ini dibutuhkan untuk menemukan suatu model yang dapat mendeskripsikan suatu class atribut itu sebagai fungsi dari input atribut. sebagai contoh: dalam pengelompokan program keluarga harapan untuk pemberian bantuan berdasarkan komponen terbanyak.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, metode ini digunakan ketika dataset yang digunakan bersifat prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode yang digunakan pada klasifikasi dan estimasi dapat digunakan untuk memprediksi suatu keadaan yang tepat.

4. Association

Association merupakan tekni yang sering disebut sebagai Market Basket Analysis. Sebuah permasalahan yang berkaitan dengan bisnis untuk menganalisa pola-pola seperti analisis tabel transaksi penjualan dan mengidentifikasi produk-produk yang sering dibeli bersamaan oleh pelanggan.

5. Clustering

Clustering juga sering di sebut dengan *segmentation* dimana metode ini sering digunakan untuk mengidentifikasi sejumlah kelompok yang bersifat alami dari suatu kasus yang didasarkan berupa sebuah kelompok atribut, dimana pengelompokan atribut tersebut memiliki kemiripan.

2.4 ALGORITMA C4.5

Salah satu solusi yang sering digunakan dalam pemecahan masalah dalam klasifikasi adalah menggunakan algoritma C4.5. layaknya teknik klasifikasi lainnya algoritma C4.5 memiliki keluaran (output), keluaran dari algoritma C4.5 berupa sebuah pohon keputusan (*decision tree*) yang dibentuk berdasarkan kriteria-kriteria pembentuk keputusan. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *records* yang lebih kecil dengan menerapkan serangkaian

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain (Berry, 2004)

Algoritma C4.5 merupakan salah satu algoritma induksi pohon keputusan yang dikembangkan dari ID3 (*Iterative Dichotomiser 3*). Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang sangat mempresentasikan aturan. Aturan yang dapat dipahami dengan mudah dengan bahasa alami. Dan pohon keputusan juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *records* pada kategori tertentu (Dicky Nofriansyah, 2014).

Secara umum untuk membangun pohon keputusan algoritma C4.5 adalah sebagai berikut :

1. Menghitung jumlah kasus dari masing-masing atribut
2. Menghitung nilai *entropy* dari masing-masing kasus
3. Menghitung nilai *gain* dari masing-masing atribut
4. Memilih atribut sebagai akar berdasarkan dari atribut yang memiliki nilai *gain* yang tertinggi.
5. Mengulangi perhitungan *entropy* dan *gain* hingga atribut tidak memiliki cabang kasus sehingga terbentuknya pohon keputusan.

2.4.1 Pembuatan pohon keputusan

Algoritma C4.5 membangun pohon keputusan dengan strategi *divide* dan *conquer*. Algoritma ini memilih pemecahan kasus-kasus yang terbaik dengan menghitung dan membandingkan *Gain Ratio*, kemudian pada node-node yang terbentuk dilevel berikutnya, algoritma *divide* dan *conquer* akan diterapkan kembali. Di dalam pembuatan pohon keputusan sebuah kasus menggunakan algoritma C4.5 ada 2 (dua) elemen yang harus dipahami yaitu:

1. *Entropy*
2. *Gain*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Entropy(S) merupakan jumlah bit yang diperkirakan akan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel *S*. *Entropy* dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai entropy maka akan semakin entropy digunakan dalam mengekstrak suatu kelas. Entropy digunakan untuk mengukur ketidakpastian *S* (Dicky Nofriansyah. Gunadi Widi Nurcahyo, 2015). adapun rumus untuk mencari nilai entropy adalah sebagai berikut:

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i \tag{2.1}$$

Keterangan:

S : himpunan kasus

n : jumlah partisi *S*

p_i : proporsi *s_i* terhadap *S*

Gain (S,A) merupakan perolehan informasi dari atribut *A* relative terhadap output *S*, perolehan informasi didapat dari output data atau variabel dependent *S* yang dikelompokkan berdasarkan atribut *A*, dinotasikan dengan *gain (S,A)* (Dicky Nofriansyah. Gunadi Widi Nurcahyo, 2015). adapun rumus untuk mencari nilai gain adalah sebagai berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|s_i|}{|S|} Entropy(S_i) \tag{2.2}$$

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut *A*

|s_i| : jumlah kasus pada partisi *k-i*

|S| : jumlah kasus dalam *S*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Sebagai contoh : kita ingin mencari apakah pemain baseball akan masuk class yes play atau no play berdasarkan data berikut:

Table 2.3 Contoh data bermain baseball

Outlook	Temperature	Humidity	Windy	Play ball
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Sumber : <http://www.cise.ufl.edu>, di akses 2020

Data-data diatas diklasifikasikan berdasarkan atribut outlook, temperature, humidity dan windy. Pembagian hasilnya ada 2, yaitu Yes dan No untuk Play ball. Dari tabel diketahui bahwa ada 14 data, 5 menyatakan No Play ball, 9 menyatakan Yes Play ball. Maka,

$$E(5,9) = -\frac{5}{9} \log_2 \frac{5}{14} - \frac{9}{14} \log_2 \frac{9}{14} = 0,940$$

Entropy untuk atribut outlook:

$$E(Outlook) = \frac{5}{14} \times E(2,3) + \frac{4}{14} \times E(4,0) + \frac{5}{14} \times E(3,2) = 0,694$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Gain untuk outlook:

$$Gain(Outlook) = E(2,3) - E(outlook) = 0,940 - 0,694 = 0,246$$

$$Gain(humadity) = 0,151$$

$$Gain(windy) = 0,048$$

$$Gain(temperature) = 0,029$$

Dengan menggunakan cara yang sama maka nilai gain dari setiap attribut dihitung. Setelah nilai gain dari setiap attribut diketahui maka yang mempunyai nilai gain terbesar yang dipilih menjadi test attribut adalah outlook.

2.4.2 Kelebihan dan Kekurangan algoritma C4.5

Algoritma C4.5 merupakan salah satu metode yang sering digunakan dalam klasifikasi data mining. layaknya teknik klasifikasi lainnya algoritma C4.5 memiliki kelebihan dan kekurangan, adapapun kelebihan algoritma C4.5 adalah sebagai berikut:

1. Daerah pengambilan lebih simpel dan spesifik.
2. Eliminasi perhitungan-perhitungan tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka sampel diuji hanya berdasarkan kriteria atau kelas tertentu.
3. Fleksibel untuk memilih fitur dari internal node yang berbeda. Sehingga dapat meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode perhitungan satu tahap yang lebih konvensional.
4. Dengan menggunakan pohon keputusan, penguji tidak perlu melakukan estimasi pada distribusi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut. Karena metode ini menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan.

Adapun kekurangan algoritma C4.5 adalah sebagai berikut:

1. Kesulitan dalam mendesain pohon keputusan yang optimal

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hasil kualitas keputusan yang didapat sangat tergantung bagaimana pohon keputusan tersebut didesain. Sehingga jika pohon keputusan yang dibuat kurang optimal, maka akan berpengaruh pada kualitas dari keputusan yang didapat.

Terjadi *overlap* terutama jika kelas-kelas kriteria yang digunakan jumlahnya sangat banyak sehingga dapat menyebabkan meningkatnya waktu pengambilan keputusan dan jumlah memori yang diperlukan.

Pengakumulasian jumlah *error* dari setiap tingkat dalam sebuah pohon keputusan yang besar.

2.5 CONFUSION MATRIX

Confusion Matrix adalah metode yang menghitung tingkat akurasi dengan menghitung jumlah prediksi benar dan salah dari sebuah metode klasifikasi berbanding dengan data sesungguhnya atau data prediksi target (Mardi, 2015). Metode ini menggunakan tabel matriks (NxN) dimana N adalah jumlah kelas. Jika *dataset* terdiri dari dua kelas, maka kelas yang datu dianggap sebagai negatif dan yang lainnya kelas negatif.

Table 2.4 Contoh Confusion Matrix

Confusion Matrix		Target			
		Positif	Negatif		
Model	Positif	A	B	Positive Predictive Value	$A/(A+B)$
	Negatif	C	D	Negative Predictive Value	$D/(C+D)$
		Sensitivity	Specifity	Accuracy= $\{(A+D)/(A+B+C)\} * 100\%$	
		$A/(A+C)$	$D/(B+D)$		



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Keterangan Label:

Accuracy : Jumlah prediksi benar (A+D)/jumlah seluruh data (A+B+C+D)

Positive predictive value : Jumlah prediksi benar *positive* (A)/jumlah target data positif (A+B)

Negative predictive value : Jumlah prediksi *benar* negatif (D)/jumlah target data negatif (B+D)

Sensitivity/Recall : jumlah prediksi benar *positive* (A)/jumlah seluruh prediksi benar (A+C)

Specificity : jumlah prediksi salah negatif (D)/jumlah seluruh prediksi salah (C+D).

2.6 PENELITIAN TERKAIT

Penelitian-penelitian terkait yang pernah dilakukan sebelumnya dan menjadi acuan penulisan tugas akhir ini adalah sebagai berikut:

No	Nama penulis	Judul	Tahun	Metode	hasil
1	irmayansyah	Penerapan algoritma C4.5 untuk klasifikasi penentuan penerimaan bantuan langsung di desa ciomas	2018	Algoritma C4.5	Hasil penelitian menunjukkan persentase kelayakan dari semua aspek penilaian yaitu 74,9% yang tergolong layak
2	Destri Arianti	Sistem diagnosa	2010	Algoritma C4.5	Dari analisa dan pengujian yang

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

		penyakit hiperkolesterol menggunakan algoritma C4.5			dilakukan sistem pendagnosa penyakit hiperkolesterol memberikan berupa penggolongan tungkatan kolestrol yaitu: hiperkilomikronemia, hiperkolesterol dll. Dari data training dan testing diperoleh <i>error</i> sebesar 5%.
3	Siti Masrifah	Evaluasi penentuan kelayakan pemberian kredit koperasi syariah menggunakan algoritma klasifikasi C4.5	2015	Algoritma C4.5	Penelitian ini menghasilkan nilai akurasi sebesar 88% dan nilai AUC sebesar 0,898 dengan tingkat diagnosa <i>good classification</i> .
4	Evicienna	algoritma C4.5 untuk prediksi hasil pemilihan legislatif DPRD DKI	2013	Algoritma C4.5	menghasilkan nilai akurasi yang tinggi yaitu sebesar 97.84% dan nilai AUC sebesar 0.970 dengan tingkat diagnosa <i>Excellent</i>



		Jakarta			<i>clasification.</i>
	Laily Hermawanti	Penerapan algoritma klasifikasi C4.5 untuk diagnosis penyakit kanker payudara	2012	Algoritma C4.5	menghasilkan nilai akurasi 94,56% dan nilai <i>area under curve</i> (AUC) sebesar 0.941 sehingga menghasilkan hasil yang akurat

Hak Cipta Dilindungi Undang-Undang

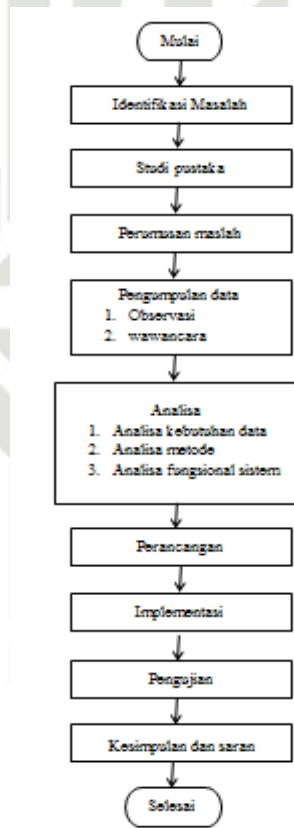
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB III METODOLOGI PENELITIAN

Metodologi penelitian merupakan tahapan peneltitan yang dilakukan secara sistematis. Metodologi penelitian pada penulisan tugas akhir ini meliputi identifikasi masalah, study pustaka, perumusan masalah, pengumpulan data, analisa, implementasi sistem, pengujian, kesimpulan dan saran seperti yang dapat dilihat dalam *flowchart* diagram seperti dibawah ini:



Gambar 3.1 Alur Metodologi Penelitian

3.1 Identifikasi Masalah

Identifikasi masalah adalah tahapan yang awal pada penelitian ini guna untuk mendapatkan gambaran dari permasalahan yang dihadapi kemudian mencari solusi dari permasalahan tersebut. Adapun masaah yang diangkat pada

penelitian ini adalah bagaimana penerapan klasifikasi data mining dengan metode algoritma C4.5 untuk menentukan kelayakan penerima bantuan program keluarga harapan (PKH).

3.2 Studi Pustaka

Pada tahap ini peneliti akan melakukan pencarian referensi, pengkajian ulang wawasan dan memperkaya dari teori yang berasal dari jurnal, *paper*, buku dan internet yang akan dikemukakan sebagai landasan dalam perhitungan dan penulisan skripsi ini. Selanjutnya, dengan adanya literatur peneliti akan mempelajari dan memahami referensi sebagai landasan dasar yang harus ditaati untuk menyelesaikan penulisan tugas akhir ini. Referensi yang digunakan pada penelitian ini meliputi studi pustaka mengenai program keluarga harapan, algoritma C4.5, pohon keputusan, k-fold cross validation, confusion matrik serta referensi terkait yang memudahkan penulisan tugas akhir ini.

3.3 Perumusan Masalah

Perumusan masalah adalah tahapan yang dilakukan setelah melakukan pencarian referensi, pengkajian ulang wawasan dan memperkaya dari teori yang berasal dari jurnal, *paper*, buku dan internet yang akan dikemukakan sebagai landasan dalam perhitungan dan penulisan. Pada tahapan ini maka ditentukanlah satu permasalahan yang akan dijadikan bahan untuk penelitian tugas akhir ini yaitu penerapan algoritma C4.5 untuk menentukan kelayakan penerima bantuan program keluarga harapan (PKH).

3.4 Pengumpulan Data

Pada tahapan ini dilakukan pengumpulan data, data yang dikumpulkan antara lain adalah data keluarga penerima manfaat program harapan berupa jumlah balita, ibu hamil, anak sd, anak smp, anak sma, pekerjaan, lanjut usia, disabilitas, status tempat tinggal, jenis dinding terluas, terima beras miskin, fasilitas jamban, tempat pembuangan akhir. Data yang digunakan merupakan data penerima manfaat program keluarga harapan kabupaten rokan hulu tahun 2017-2019. Teknik pengumpulan data dalam penelitian ini antar lain:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Observasi langsung, peneliti mengamati langsung keadaan masyarakat penerima bantuan program keluarga harapan
2. Wawancara, peneliti melakukan tanya jawab dengan pihak pendamping desa program keluarga harapan sebagai pihak yang terkait langsung dalam penelitian guna memperoleh data tentang kriteria yang diperlukan untuk menentukan keluarga penerima bantuan program keluarga harapan dan pihak dinas sosial selaku pihak yang akan memberikan data yang diperlukan dalam penelitian.

3.5 Analisa

Pada penelitian tugas akhir ini dilakukan tiga tahapan analisa yaitu analisa kebutuhan data dan analisa metode. Berikut adalah penjelesan dari masing-masing tahapan .

3.5.1 Analisa Kebutuhan Data

Pada tahapan ini dilakukan analisa pada beberapa kriteria yang berhubungan dengan penelitian ini. Kriteria yang terkait dengan bantuan program keluarga harapan jumlah balita, ibu hamil, anak sd, anak smp, anak sma, pekerjaan, lanjut usia, disabilitas, status tempat tinggal, jenis dinding terluas, terima beras miskin, fasilitas jamban, tempat pembuangan akhir. komponen dan kriteria penerima bantuan program keluarga harapan yang meliputi indeks kesehatan, pendidikan, dan kesejahteraan sosial. Setelah kriteria ditentukan dan data yang diperlukan telah dikumpulkan maka dilakukan proses *data selection*, *pre-processing*, *transformation* dan data mining. Pada proses *data selection* akan dilakukan pemilihan variabel serta attribut yang akan digunakan dalam penelitian ini. Pada proses *pre-processing* akan dilakukan tahapan guna untuk mengatasi masalah data yang hilang (*missing value*), data ganda, dan memperbaiki kesalahan yang terdapat pada data yang tidak konsisten. Pada proses transformation adalah tahapan dalam proses *coding* yaitu melakukan tranformasi pada data yang telah dipilih sesuai kebutuhan untuk dilanjutkan pada proses data mining. pada proses data mining dilakukan penerapan algoritma C4.5 untuk klasifikasi data yang telah melalui tahapan-tahapan yang telah dilakukan sebelumnya.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.5.2 Analisa metode

Pada tahapan analisa metode dilakukan analisa terhadap metode yang digunakan dalam penelitian ini yaitu algoritma C4.5. analisa algoritma C4.5 yaitu membangun pohon keputusan berdasarkan nilai gain tertinggi. tahapan penerapan algoritma C4.5 adalah sebagai berikut:

1. Menghitung jumlah kasus dari masing-masing atribut
2. Menghitung nilai *entropy* dari masing-masing kasus dengan menggunakan persamaan (2.1).
3. Menghitung nilai *gain* dari masing-masing atribut dengan menggunakan persamaan (2.2).
4. Memilih atribut sebagai akar berdasarkan dari atribut yang memiliki nilai *gain* yang tertinggi.
5. Mengulangi perhitungan *entropy* dan *gain* hingga atribut tidak memiliki cabang kasus sehingga terbentuknya pohon keputusan.

3.6 Implementasi

Pada tahapan ini dilakukan untuk menjalankan program yang telah dirancang dan dibuat sebelumnya. Kemudian dilakukan pengecekan sistem apakah program yang telah dibuat telah sesuai dengan apa yang telah direncanakan sebelumnya dan telah mencapai tujuan yang diinginkan pada penelitian ini. Berikut adalah perangkat yang digunakan dalam penelitian ini:

1. Perangkat keras (*Hardware*)
 - a. Processor : Intel Corei3
 - b. RAM : 4 GB
 - c. *Harddisk* : 500 GB
2. Perangkat Lunak (*Software*)
 - a. Operating System : Microsoft Windows 8.1 64-bit
 - b. Library : *pandas, numpy,pydotplus, Ipython. display.*
 - c. Browser : Google Chrome
 - d. Bahasa Pemrograman : Python



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.7 Pengujian

Pada tahap ini pengujian ini dilakukan untuk mengetahui apakah perhitungan algoritma sudah tepat, Pengujian yang dilakukan dengan program yang telah dibuat menggunakan metode *Coinfusion Matrix* dengan tujuan menghitung tingkat akurasi algoritma C4.5 dalam menentukan kelayakan penerima bantuan program keluarga harapan (PKH) .

3.8 Kesimpulan dan Saran

Pada tahap ini berisi kesimpulan dan saran. Bagian kesimpulan merupakan penentuan hasil pengujian yang telah dilakukan, apakah hasil dari penelitian yang telah dilakukan sesuai dengan apa yang diharapkan atau tidak dan tingkat keberhasilan pada suatu penelitian. Pada bagian saran berisi kemungkinan pengembangan yang akan dilakukan terhadap penelitian ini dan tentang perbaikan-perbaikan yang diharapkan dapat menghasilkan penelitian yang lebih baik sehingga menjadi referensi yang akurat untuk penelitian selanjutnya.



UIN SUSKA RIAU

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau



UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB VI PENUTUP

4.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, maka diperoleh beberapa kesimpulan dari penerapan algoritma *C4.5* untuk mengklasifikasi penerima bantuan program keluarga harapan (PKH) adalah sebagai berikut:

1. Penerapan algoritma *C4.5* untuk mengklasifikasi penerima bantuan program keluarga harapan (PKH) telah berhasil dilakukan.
2. Pohon keputusan yang dihasilkan dengan algoritma *C4.5* menjadikan atribut anak smp memiliki nilai Gain tertinggi sebagai acuan penentuan kelayakan penerima bantuan.
3. Pengujian algoritma *C4.5* menggunakan *confusion matrix* memperoleh akurasi terbaik dengan perbandingan 90% data latih dan 10% data uji sebesar 86,37%.
4. Berdasarkan hasil pengujian tersebut algoritma *C4.5* termasuk pada range *Good Clasification*.

4.2 Saran

Adapun saran untuk pengembangan penelitian ini di masa yang akan datang terkait penelitian ini yaitu:

1. Menggunakan lebih banyak atribut yang termasuk dalam kriteria penerima bantuan program keluarga harapan agar mendapatkan akurasi yang lebih tinggi lagi.
2. Menambah jumlah data menjadi lebih banyak sehingga dapat mempengaruhi tingkat akurasi pada proses klasifikasi.
3. Berdasarkan penelitian ini dibuatnya software berbasis web atau android yang dapat menentukan kelayakan penerima bantuan program keluarga harapan.



4. Menggunakan metode klasifikasi data mining lainnya sebagai bahan perbandingan dalam menentukan kelayakan penerima bantuan program keluarga harapan.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



UIN SUSKA RIAU

Daftar pustaka

- Anzir, Q. (2019). penerapan metode probabilistic neural networ untuk klasifikasi program keluarga harapan kota pekanbaru. *penerapan metode pnn*, 15.
- Berry, G. S. (2004). *Data Mining Techniques*. willey publishing, Inc.
- Dicky Nofriansyah, S. M. (2014). *Konsep Data Mining vs Sistem Pendukung Keputusan*. yogyakarta: Deepublish.
- Dicky Nofriansyah. Gunadi Widi Nurcahyo, M. (2015). *Algoritma Data Mining dan Pengujian*. yogyakarta: CV BUDI UTAMA.
- fitriani, e. (2019). Perbandingan algoritma C4.5 dan naïve bayes untuk menentukan kelayakan penerima bantuan program keluarga harapan (PKH). *perbandingan algoritma C4.5* , 15.
- Han, J. K. (2012). *Data Mining: Concept and Techniques, third edition*. waltham: morgan kaufmann .
- kemosos. (2018, september 1). *program keluarga harapan*. Retrieved april 15, 2020, from program keluarga harapan kementrian sosial republik indonesia: <https://pkh.kemosos.go.id/?pg=tentangpkh-1>
- Kuncoro, M. (2006). *ekonomi pembangunan: teori , masalah, dan kebijakan*. yogyakarta: lembaga penerbit fakultas.
- Musrini, d. (2009). *Algoritma data mining*. yogyakarta: andi offset.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Elarose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, Inc.

Elroy, B. (2001). *Machine Learning*. Berkeley: Universitas California.

Elwan, J. L. (2003). Comparing Naive Bayes, Decision Trees, and SVM with accuracy. *The Third IEEE International Conference on Data Mining*, 11-14.

Mardi. (2015). Data Mining: Klasifikasi menggunakan algoritma C4.5. *Jurnal edik informatika*, 213-219.

Polamuri, S. (2017). *How Random Forest Works in Machine Learning*. Data Aspirant.

Sosial, K. (2019). *Petunjuk Teknis Penyaluran Bantuan Tunai Program Keluarga Harapan*. Jakarta: Kementerian Sosial.

Turban, E. (2005). *Decision Support Systems and Intelligent Systems*. Yogyakarta: Andi Offset.

Wicaksana., d. a. (2013). *Belajar Data Mining Dengan Rapid Miner*. Jakarta.

[BPS] Badan Pusat Statistik. (2019). Berita resmi statistik. *Bps.Go.Id*.

Anjarwati, S., & Farahdibah. (2017). Sistem Pendukung Keputusan Seleksi Calon Penerima Bantuan Program Keluarga Harapan (PKH) dengan Metode AHP dan Promethee (Studi Kasus pada Kelurahan Kudaile Slawi). *Surya Informatika*, 4(1), 1-14.

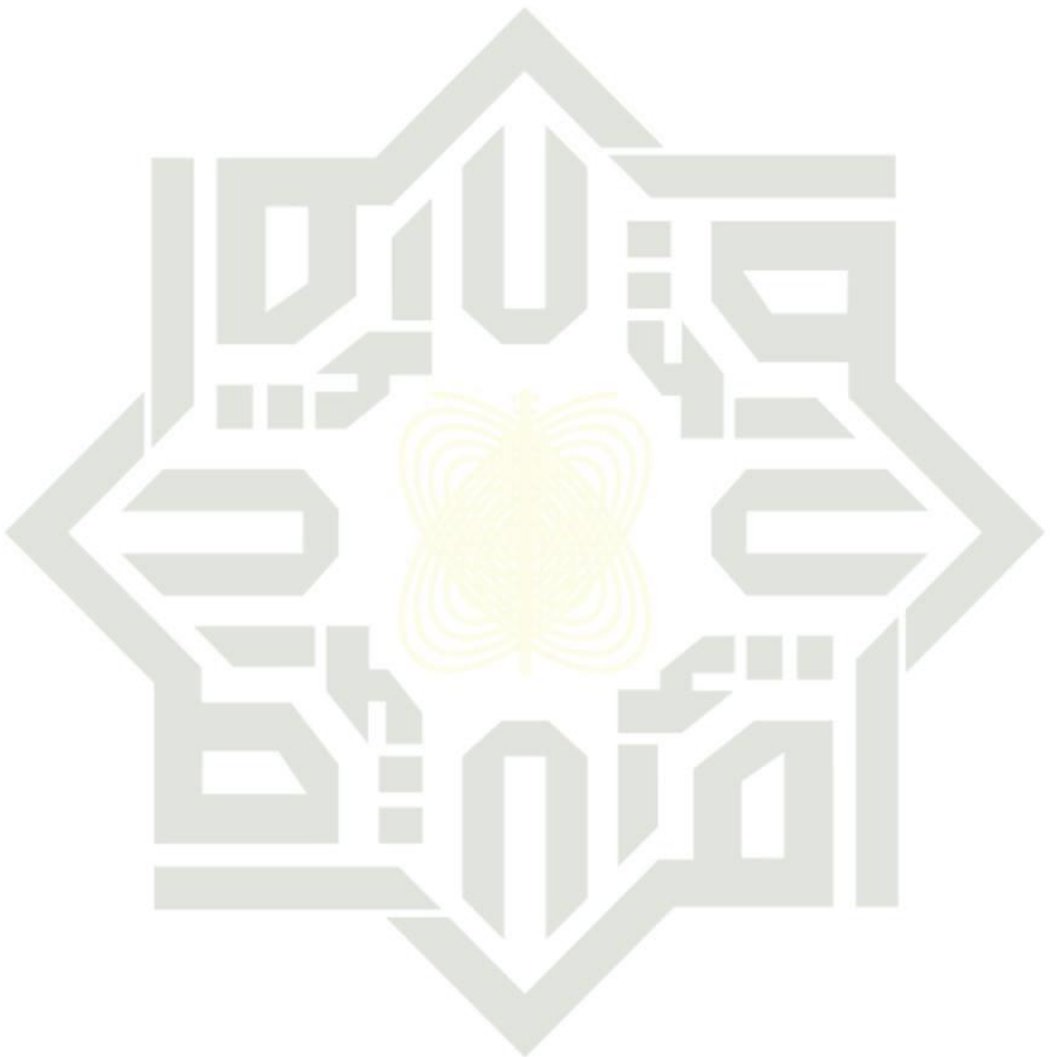
Musrini, & Emha, T. (2015). Definisi Data Mining. *Data Mining*.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Zurwati, N. (2008). Kemiskinan: Model Pengukuran, Permasalahan dan Alternatif Kebijakan. *Jurnal Kependudukan Padjadjaran*.



UIN SUSKA RIAU