

BAB II KERANGKA TEORITIS

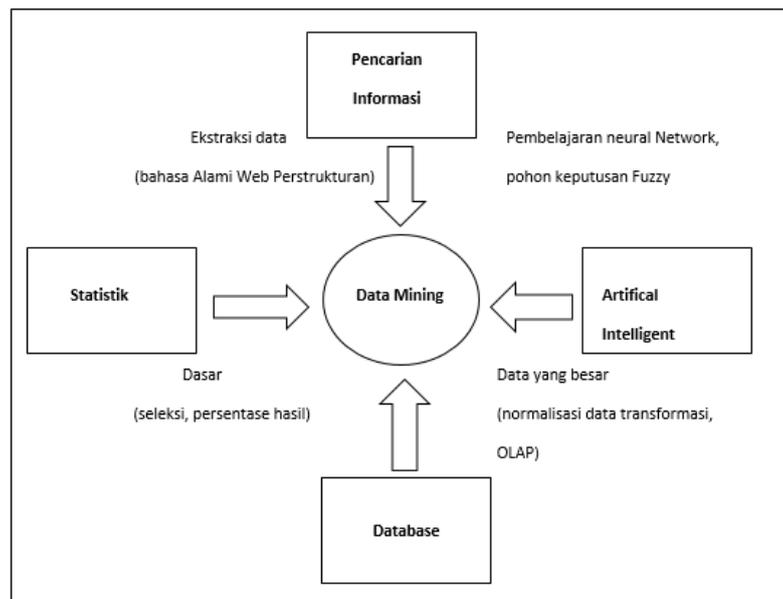
A. Landasan Teori

1. Data Mining

Data Mining menurut (Pramudiono, 2006 : 6) merupakan analisis otomatis dari data yang berjumlah banyak atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang umumnya tidak disadari keberadaannya.

Berdasarkan (Turban,dkk. 2005 : 5) data mining ialah suatu istilah untuk menguraikan penemuan pengetahuan pada database. Data mining merupakan proses yang menggunakan teknik static, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat serta pengetahuan yang terkait dari berbagai database besar.

Ada juga menurut Garnet Group, data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dengan sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2005 : 5).



Gambar 2.1. Bidang Ilmu Data Mining

(Sumber : Nofriansyah, 2014 : 6)

Data mining secara umum adalah kegiatan pencarian (*discovery*) secara berulang (*iterative*) dan intensif yang bertujuan untuk mengekstrak pengetahuan dari kumpulan data yang tadinya data yang tadinya tidak/belum membawa arti penting. Pengetahuan yang dimaksud dapat berupa pattern/pola, hubungan, perubahan, anomaly, struktur, rumus, aturan, ataupun model yang muncul dari data. Hasil yang didapatkan harus valid, berguna dan mudah dimengerti.

2. Business Process Model and Notation (BPMN)

Dari Paul Harmon, dalam bukunya “Business Process Change” (2003), definisi proses bisnis artinya serangkaian aktifitas yang dilakukan oleh suatu usaha dimana mencakup inisiasi input, transformasi dari suatu informasi, dan menghasilkan output. Hasil tadi dapat bernilai bagi pelanggan bisnis atau market, dapat juga bernilai bagi proses yang lain (dalam organisasi). Suatu proses bisnis dapat dipecah menjadi beberapa sub-proses yang masing-masing memiliki atribut sendiri yang berkontribusi untuk mencapai tujuan dari proses induknya. Sub-proses dapat dipecah lagi menjadi aktifitas, yaitu sub-proses terkecil yang dapat terdiri dari satu atau lebih langkah (steps) yang harus dicantumkan dalam proses bisnis.

3. Unified Modeling Language (UML)

Unified Modeling Language (UML) merupakan bahasa yang telah menjadi patokan dalam industri untuk penggambaran, merangka dan pengarsipan sistem peranti lunak, UML menawarkan sebuah standar untuk merancang model sebuah sistem (Sugiarti, 2018 : 101).

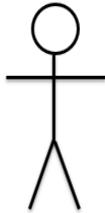
Arsitektur perlu didefinisikan dengan jelas, agar *bug* mudah ditemukan dan diperbaiki. Keuntungan lain dari perencanaan arsitektur adalah dimungkinkannya pengguna kembali modul atau komponen untuk aplikasi peranti lunak lain yang membutuhkan fungsionalitas yang sama (Sugiarti, 2013 : 100).

Ada beberapa diagram dasar dalam UML, yaitu :

a. Model Usecase Diagram

Menurut Sugiarti (2018 : 111) diagram usecase adalah penjelasan sebuah hubungan antara satu atau lebih aktor dengan sistem yang akan dibuat, diagram usecase digunakan bagi memahami manfaat yang ada di dalam sistem dan yang berwenang menggunakan manfaat tersebut. Dalam bagan usecase ada simbol – simbol yang sudah ditentukan :

Tabel 2.1. Simbol Usecase

Simbol	Nama Simbol	Keterangan
	Usecase	Manfaat yang disediakan sistem sebagai elemen yang saling bertukar pesan antar unsur atau aktor; umumnya dinyatakan dengan memakai kata kerja di awal nama usecase.
	Aktor	Orang, proses, atau sistem lain yang berhubungan dengan sistem informasi yang akan dilakukan di luar sistem informasi yang akan dibuat itu sendiri. Jadi walaupun simbol dari aktor adalah gambar orang, tetapi aktor belum tentu merupakan orang; umumnya dijelaskan menggunakan kata benda di awal nama aktor.
	Asosiasi / association	Kontak antara aktor dan usecase yang terlibat pada usecase atau usecase memiliki hubungan dengan aktor.
 <<extend>>	Extend	Perhubungan usecase ditambah ke sebuah usecase dimana yang ditambahkan dapat berdiri sendiri walaupun tanpa usecase tambahan.
 <<include>>	Include	Perhubungan usecase ekstra ke sebuah usecase di mana usecase yang ditambahkan membutuhkan usecase ini untuk melakukan manfaat atau sebagai ketentuan dilakukan usecase ini.

Simbol	Nama Simbol	Keterangan
	<i>GENERALISASI</i>	Hubungan generalisasi dan spesialisasi (umum-khusus) dua buah use case dimana fungsi yang satu adalah fungsi yang lebih umum dari lainnya.

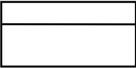
(Sumber : Sugiarti, 2018 : 111)

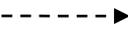
b. Diagram Kelas (*Class Diagram*)

Diagram kelas atau *class diagram* melukiskan tata sistem dari segi pendefinisian *class – class* yang akan dikerjakan untuk membuat sistem. *Class diagram* menjelaskan struktur dan uraian *class*, *package*, dan objek beserta hubungan satu sama lain seperti *containment*, pewarisan, asosiasi, dan lain – lain.

Berikut simbol – simbol yang terdapat pada *class diagram* :

Tabel 2.2. Simbol *Class Diagram*

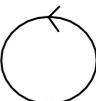
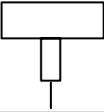
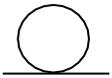
Simbol	Nama Simbol	Keterangan
	<i>Generalization</i>	Hubungan dimana objek anak (descendent) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk (ancestor).
	<i>Nary Association</i>	Upaya untuk menghindari asosiasi dengan lebih dari 2 objek.
	<i>Class</i>	Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama.
	<i>Collaboration</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu actor.
	<i>Realization</i>	Operasi yang benar-benar dilakukan oleh suatu objek

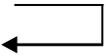
Simbol	Nama Simbol	Keterangan
	<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (independent) akan mempengaruhi elemen yang bergantung padanya elemen yang tidak mandiri.
	Association	Apa yang menghubungkan antara objek satu dengan obyek lainnya.

c. Diagram Interaksi (*Sequence Diagram*)

Diagram sekuens atau *sequence diagram* adalah gambaran *behavior objek* pada usecase dan message yang dikirimkan dan diterima antarobjek. Untuk menggambarkan diagram sekuens jadi patut diketahui objek – objek yang terlibat dalam sebuah usecase beserta metode – metode kepunyaan class yang diinstansiasi menjadi objek itu (Sugiarti, 2018).

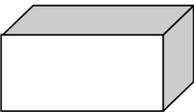
Tabel 2.3. Diagram Sequence

Simbol	Nama	Keterangan
	<i>Control</i>	Bagian menata gerakan dari informasi untuk sebuah skenario. Bahan ini umumnya sifat dan sifat bisnis.
	<i>Actor</i>	Menggambarkan entitas yang berada di luar sistem dan berhubungan di luar sistem.
	<i>General</i>	Menggambarkan entitas tunggal dalam sequence.
	<i>Boundary</i>	Bersifat batas dari sistem, seperti user interface dan perangkat yang berhubungan dengan yang lain.
	<i>Entitas</i>	Bagian yang bertugas menyimpan informasi.

Simbol	Nama	Keterangan
	<i>Lifeline</i>	Menyambungkan objek selama sequence (<i>message</i> dikirim atau diterima).
	<i>Activation</i>	Suatu titik dimana sebuah objek mulai ikut serta dalam sebuah sequence yang menunjukkan sebuah objek mengirim atau menerima objek.
	<i>Message Entry</i>	Berfungsi untuk melukiskan pesan antar objek yang menunjukkan urutan kejadian.
	<i>Message to Self</i>	Simbol ini melukiskan pesan/hubungan objek, yang menunjukkan urutan kejadian yang terjadi.
	<i>Message Return</i>	Melukiskan hasil dari pengiriman pesan yang digambarkan dengan arah dari kanan ke kiri.

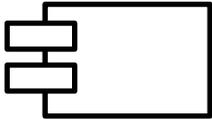
d. Deployment Diagram

Tabel 2.4. Deployment Diagram

Simbol	Nama	Keterangan
	<i>Package</i>	Berupa satu atau lebih tempat dari 1 atau lebih cara.
	<i>Node</i>	Biasanya mengacu pada perangkat keras (<i>hardware</i>), perangkat lunak yang tidak dibuat sendiri (<i>software</i>), jika didalam node disertakan unsur untuk menstabilkan rancangan.
	Kebergantungan/ <i>dependency</i>	Kebergantungan antar node, arah panah mengarah pada node.

e. Komponen Diagram (*Component Diagram*)

Tabel 2.5. Component Diagram

Simbol	Nama	Keterangan
	Komponen	Komponen apa saja yang ada pada sistem
	Kebergantungan/ <i>dependency</i>	Penghubung yang menuju pada node yang digunakan.
	Link	Penghubung antara node satu dengan node lainnya.
	<i>Package</i>	Kumpulan dari beberapa komponen

4. Klasifikasi

Klasifikasi merupakan sebuah proses pembelajaran (*training*) suatu pokok tujuan (*target*) yang digunakan tiap melukiskan tiap kumpulan atribut suatu objek dari satu ke label kelas tertentu yang dijelaskan sebelumnya. Teknik klasifikasi ini cocok digunakan di dalam menjelaskan data-set dengan tipe data dari suatu kumpulan data yaitu biner atau nominal. Adapun kekurangan dari teknik ini yaitu ketidak persis untuk kumpulan data ordinal karena penghampiran yang digunakan sebagai sugestif dalam jenis data (Suyanto, 2017).

Algoritma adalah model untuk memprediksi kelas atau kategori untuk satu objek data. Misalnya, proses untuk menentukan apakah suatu email yang masuk adalah *spam* atau bukan dengan teknik klasifier biner. Ada dua bentuk tugas klasifikasi. Yang pertama adalah klasifikasi biner, dimana tujuannya untuk memprediksi dengan satu objek terdiri dari dua kelas. Yang kedua adalah klasifikasi multi kelas, yang digunakan untuk memprediksi satu dari banyak kelas. Output dari algoritma klasifikasi disebut *classifier*, yang dapat digunakan untuk memprediksi label sebuah instance (*unlabeled*) data baru (Suyanto, 2017).

Decision tree merupakan salah satu gaya klasifikasi yang populer dan banyak digunakan secara efektif. Salah satu metode *decision tree* yang

sangat populer adalah *Iterative Dychotomizer version 3* (ID3). Dua versi lain yang populer ialah C4.5 dan ASSISTANT.

Ada beberapa ukuran yang dapat dipakai untuk menilai atau penilaian model klasifikasi adalah : *accuracy* atau tingkat pengenalan, *error rate* atau tingkat kesalahan atau kekeliruan klasifikasi, *recall* atau *sesitivity* atau *true positive rate*, *specificity* atau *true negative rate*, *precision*, *F-measure* atau *F-score* atau rata – rata kesesuaian dari *precision* dan *recall* (Suyanto, 2018).

5. System Development Life Cycle (SDLC)

Bagi Raymond McLeod (2007), sesungguhnya pendekatan sistem membuat sebuah metodologi. Metodologi adalah satu cara yang direkomendasikan dalam melakukan sesuatu. Pendekatan sistem adalah metodologi dasar dalam memecahkan semua bentuk masalah. Peredaran hidup pengembangan sistem (*System Development Life Cycle*–SDLC) adalah aplikasi dari ancangan sistem bagi peningkatan suatu sistem informasi.

Terdapat sejumlah tahapan pekerjaan pengembangan yang perlu dilakukan jika suatu proyek ingin memiliki kemungkinan berhasil yang besar. Tahapan tersebut adalah:

- a. Perencanaan
- b. Analisis
- c. Desain
- d. Implementasi
- e. Penggunaan.

Proyek dan sumber daya yang dibutuhkan untuk melakukan pekerjaan direncanakan lalu disatukan. Sistem yang ada juga dianalisis untuk memahami masalah dan menentukan persyaratan fungsional dari sistem yang baru. Sistem baru ini kemudian dirancang dan diterapkan. Setelah diterapkan, sistem lalu digunakan, idealnya untuk jangka waktu yang lama.

Karena operasi di atas mengikuti satu pola yang teratur dan dilakukan dengan cara dari atas ke bawah, SDLC tradisional sering disebut sebagai pendekatan air terjun (*waterfall approach*). Kegiatan ini memiliki gerakan satu arah menuju ke penyelesaian proyek.

Mudah untuk kita melihat SDLC tradisional dapat dikatakan sebagai suatu aplikasi dari pendekatan sistem. Masalah akan dijelaskan dalam tahapan perencanaan dan analisis. Penyelesaian pilihan dikenali dan ditingkatkan

dalam tahap desain. Lalu, solusi yang terbaik diimplementasikan dan digunakan. Selama tahap penggunaan, umpan balik dikumpulkan untuk melihat seberapa baik sistem mampu menyelesaikan masalah yang telah ditentukan (Raymond McLeod, 2007:188).

6. Bahasa Pemrograman

a. Pengertian PHP

Menurut Anhar (2010 : 3), menjelaskan bahwa PHP singkatan dari *Hypertext Preprocessor* yaitu bahasa pemrograman web *server-side* yang bersifat *open source* atau diperuntukan pengembangan web dan dapat ditanamkan pada sebuah skripsi HTML berdasarkan kode-kode (*script*) yang dipakai untuk mengolah suatu data dan mengirimkannya kembali ke web browser menjadi kode HTML.

b. Pengertian Hypertext Markup Language (HTML)

Menurut Simarmata (2010 : 52), menjelaskan HTML adalah bahasa *markup* untuk menyebarkan informasi pada web. Bahasa pemrograman patokan yang digunakan untuk membuat sebuah halaman web, yang kemudian dapat diakses untuk menampilkan berbagai informasi di dalam sebuah penjelajah *web internet (browser)*.

Informasi yang tertata di dalam arsip atau sekumpulan dokumen. Walaupun HTML tidak dengan mudah dapat dipahami kebanyakan orang, ketika diterbitkan penggunaannya menjadi jelas.

c. Pengertian MySQL

Menurut Kustiyahningsih (2011 : 145-147), menjelaskan My SQL merupakan sebuah basis data yang mengandung satu atau jumlah tabel. Tabel terdiri atas sejumlah baris dan setiap baris mengandung satu atau sejumlah tabel. Tabel terdiri atas sejumlah baris dan setiap baris mengandung satu atau sejumlah tabel.

Tipe data My SQL adalah data yang terdapat dalam sebuah tabel berupa *field-field* yang berisi nilai dari data tersebut. Nilai data dalam *field* memiliki tipe berbeda.

d. Pengertian Database

Dari Mustakini (2009 : 46), menjelaskan bahwa database adalah kumpulan dari data yang saling berhubungan satu dengan yang lainnya,

tersimpan di perangkat keras komputer dan digunakan perangkat lunak untuk memanipulasi.

7. Web Server (Apache)

Menurut Kurniawan (2008: 2), menjelaskan bahwa *web server* adalah sebuah perangkat lunak server yang berfungsi menerima permintaan HTTP atau HTTPS dari klien yang dikenal dengan web browser dan mengirimkan kembali hasilnya dalam bentuk halaman-halaman web yang umumnya berbentuk halaman HTML. Web server yang terkenal diantaranya ialah Apache dan *microsoft internet information service (IIS)*.

8. Metode Prototype

Menurut (Ogedebe, 2012) bahwa *prototyping* adalah gaya pengembangan perangkat lunak, yang berupa model fisik kerja sistem dan berfungsi sebagai versi awal dari sistem. Dengan cara *prototyping* ini akan menghasilkan *prototype* sistem sebagai penghubung pengembang dan pengguna agar dapat berhubungan dalam proses kegiatan pengembangan sistem informasi. Agar proses pembuatan *prototype* ini selesai dengan baik adalah dengan menjelaskan kebutuhan awal. *Prototype* akan dihilangkan atau ditambahkan pada bagiannya sehingga sesuai dengan perencanaan dan analisis yang dilakukan oleh pengembang sampai dengan uji coba dilakukan secara beriringan dengan proses pengembangan. Ada 4 metodologi *prototyping* yang paling utama yaitu :

Illustrative, menghasilkan contoh laporan dan tampilan layar.

Simulated, mensimulasikan beberapa alur kerja sistem tetapi tidak menggunakan data real.

Functional, mencoba beberapa alur sistem yang sebenarnya dan menggunakan data real.

Evolutionary, menghasilkan cara yang menjadi bagian dari operasional sistem.

Dibuatnya sebuah *prototyping* bagi pengembang sistem bertujuan untuk mengumpulkan informasi dari pengguna sehingga pengguna dapat berinteraksi dengan cara *prototype* yang dikembangkan, sebab *prototype* menggambarkan versi awal dari sistem untuk kelanjutan sistem sebenarnya yang lebih besar.

(Ogedebe, 2012), menekankan telah ditemukan dalam analisis dan rancangan sistem, terutama untuk proses transaksi, dialog yang ditunjukkan lebih mudah dimengerti. Semakin besar hubungan antara komputer dan pengguna, besar pula manfaat yang didapat ketika proses pengembangan sistem informasi akan lebih cepat dan membuat pengguna akan lebih berhubungan dalam proses pengembangannya.

Prototyping dapat diterapkan pada pengembangan sistem kecil atau besar dengan harapan agar proses pengembangan dapat berjalan dengan baik, tertata serta dapat selesai tepat waktu. Keterlibatan pengguna secara penuh ketika prototype terbentuk akan menguntungkan seluruh pihak yang terlibat, bagi pimpinan, pengguna sendiri serta pengembang sistem.

Manfaat lainnya dari penggunaan prototyping adalah :

1. Mewujudkan sistem sebenarnya dalam bentuk replika sistem yang akan digunakan, menadahkan masukan dari pengguna untuk kesempurnaan sistem.
2. Pengguna akan lebih siap menerima setiap perubahan sistem yang berkembang sesuai dengan berjalannya prototype hingga dengan akhir pengembangan yang akan digunakan nantinya.
3. Prototype dapat ditambah maupun dikurangi sesuai berjalannya proses pengembangan. Kemajuan tahap dapat diikuti oleh pengguna.
4. Penghematan sumber daya dan waktu dalam menghasilkan produk yang lebih baik dan tepat untuk pengguna.

B. Penerima Bantuan Program Keluarga Harapan (PKH)

Klasifikasi merupakan cara dalam data mining untuk membagi data berdasarkan ketertarikan data pada data sampel.

PKH singkatan dari Program Keluarga Harapan adalah bantuan pemerintah dalam cara mengurangi kemiskinan. Menurut Kementerian Sosial Negara Republik Indonesia yang di lansir pada situs web resmi Kementerian Sosial menjelaskan Program Keluarga Harapan yang selanjutnya disebut PKH ialah program pemberian bantuan sosial bersyarat untuk Keluarga Miskin (KM) yang ditetapkan sebagai keluarga penerima manfaat PKH. Sebagai upaya percepatan pengendalian kemiskinan. Klasifikasi di bantuan PKH dapat memberikan hasil klasifikasi yang tepat sasaran bagi penerimanya. Ada beberapa jenis penerima bantuan PKH dibutuhkan klasifikasi bantuan PKH untuk anak usia sekolah. Kategori anak usia sekolah adalah SD, SMP dan SMA.

C. Algoritma C4.5

Algoritma C4.5 adalah algoritma yang dipakai untuk membuat pohon keputusan. Di dalam algoritma C4.5 ini, pohon – pohon keputusan yang dibuat berdasarkan kriteria – kriteria pembentuk keputusan.

Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi kumpulan *record* yang lebih kecil dengan memakai serangkaian aturan keputusan. Dengan masing – masing rangkaian donasi, anggota himpunan hasil menjadi mirip satu dengan yang lain (Berry dan Linoff, 2004).

Untuk penyelesaian kasus di dalam algoritma C4.5 ada beberapa unsur yang diketahui, yaitu :

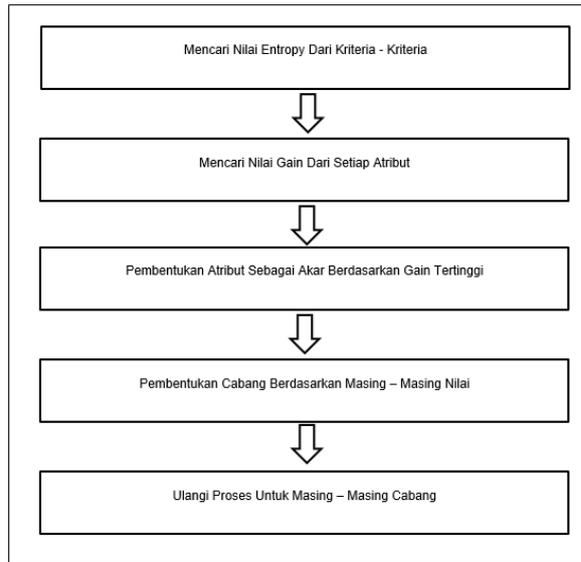
1. Entropy (S)

Merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S. Entropy dapat disebut sebagai kebutuhan bit untuk menyampaikan suatu kelas. Semakin kecil nilai Entropy maka akan semakin Entropy digunakan dalam mengekstrak suatu kelas. Entropy dipakai untuk mengukur ketidak aslian S.

2. Gain (S, A)

Merupakan perolehan informasi dari atribut A relatif terhadap output data S. Perolehan informasi didapat dari output data atau variabel dependent S yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (S, A).

Adapun langkah untuk penyelesaian algoritma C4.5 terlihat pada peredaran dibawah ini :



Gambar 2.2. Alur Penyelesaian Algoritma C4.5

(Sumber : Nofriansyah dan Nurcahyo, 2015 : 19)

Contoh Kasus

Seperti yang ditulis oleh Nofriansyah dan Nurcahyo contoh masalah yang akan di analisis adalah untuk mengklasifikasi calon pendaftar di suatu STMIK Triguna dalam hal pemilihan program studi khususnya : Sistem Komputer atau Sistem Informasi. Adapun data yang dipakai dalam membuat pohon keputusan untuk menganalisis minat calon mahasiswa baru untuk mendaftar ke STMIK Triguna berdasarkan program studi strata 1 adalah nama calon mahasiswa, asal sekolah, jenis kelamin, hobi. Data selanjutnya akan dilakukan pra-proses untuk menghasilkan data kasus yang siap dibuat untuk menjadi sebuah pohon keputusan.

Data yang tidak lengkap disebabkan karena ada data yang kosong atau atribut yang salah. Begitu juga dengan data minat calon mahasiswa baru yang mendaftar ke STMIK Triguna berdasarkan program studi strata, ada beberapa atribut yang tidak perlu sehingga proses Data Preprocessing perlu dilakukan sehingga database sesuai dengan ketentuan yang diperlukan.

Data Preprocessing merupakan hal yang penting dalam proses data mining, hal yang termasuk yaitu :

1. Data Selection

Data minat calon mahasiswa/i baru yang mendaftar ke STMIK Triguna berdasarkan program studi strata 1 tersebut akan menjadi data kasus dalam

proses operasional data mining. Dari data yang ada, kolom yang diambil atribut keputusan adalah hasil, sedangkan kolom yang diambil sebagai atribut penentuan dalam pembentukan pohon keputusan adalah nama mahasiswa, minat calon mahasiswa, asal sekolah, jenis kelamin, dan hobi.

2. Data Cleaning

Data cleaning disusun untuk menambah isi atribut yang hilang atau kosong dan merubah data yang tidak konsisten.

3. Data Transformation

Dalam proses ini, data dipindahkan ke dalam bentuk yang sesuai untuk proses data mining.

4. Data Reduction

Reduksi data dilakukan dengan menghilangkan atribut yang tidak diperlukan sehingga ukuran dari database menjadi kecil dan hanya menyertakan atribut yang diperlukan dalam proses data mining, karena lebih efisien terhadap data yang lebih kecil.

Masalah klasifikasi berakhir dengan dihasilkan sebuah pengetahuan yang digambarkan dalam bentuk diagram yang biasa disebut pohon keputusan (*decision tree*). Data berikut ini dipergunakan untuk data latihan. Data selengkapnya tampak pada tabel dibawah ini :

Tabel 2.6. Data Latihan

No	Nama Mahasiswa	Calon Mahasiswa	Asal Sekolah	Jenis Kelamin	Hobi	Hasil
1	Novita Dewi B.B	Hardware	SMK Komputer	Laki – laki	Non	SK
2	Ahmad Riyandi	Hardware	SMK Komputer	Laki – laki	IT	SK
3	Reza Ardiansyah	Umum	SMK Komputer	Laki – laki	Non	SI
4	Gafar Dwi Satrio	Software	SMA UMUM	Laki – laki	Non	SI
5	Nur Azizah Dalimunthe	Software	SMK TEKNIK	Perempuan	Non	SI
6	Ray Ishak Permana B	Software	SMK TEKNIK	Perempuan	IT	SI
7	M. Rizky Fadly	Umum	SMK TEKNIK	Perempuan	IT	SI
8	Zulfikar Ali	Hardware	SMA UMUM	Laki – laki	Non	SK
9	Putra Mustaqim	Hardware	SMK TEKNIK	Perempuan	Non	SI

No	Nama Mahasiswa	Calon Mahasiswa	Asal Sekolah	Jenis Kelamin	Hobi	Hasil
10	Debby L. Simatupang	Software	SMA UMUM	Perempuan	Non	SI
11	Daniel A. Sihombing	Hardware	SMA UMUM	Perempuan	IT	SI
12	Asri Anzani Br. Tarigan	Umum	SMA UMUM	Laki - laki	IT	SI
13	Abdul Alim	Umum	SMK Komputer	Perempuan	Non	SI
14	Akbar Widianara	Software	SMA UMUM	Laki - laki	IT	SK

(Sumber : Nofriansyah dan Nurcahyo, 2015 : 21)

Keterangan :

Untuk Asal Sekolah yang disebut SMK Komputer yaitu berasal dari Jurusan Teknik Komputer dan Jaringan, Multimedia, dan Rekayasa Perangkat Lunak sedangkan yang dikatakan sebagai sekolah umum yaitu Sekolah Menengah Atas yang terdiri dari jurusan IPA dan IPS dan yang dimaksud dengan SMK Teknik adalah yang berasal dari jurusan baik Teknik Elektro, Teknik Mesin dan lain – lain. Hasil Sistem Informasi dan SK merupakan Nilai Atribut Hasil Sistem Komputer.

Setelah kita memperoleh data minat calon mahasiswa/i baru yang tercantum pada tabel 2.4. Langkah selanjutnya adalah menentukan Entropy dan Gain :

1. Nilai Entropy

a. **Entropy Total** = $Entropy (S) = \sum_{i=1}^n - p_i * \log_2 p_i$

$$Entropy \text{ Total} = ((-4/14 * \log_2 (4/14)) + (-10/14 * \log_2 (10/14))) \\ = 0.863120569$$

b. **Entropy Minat Calon Mahasiswa**

- Nilai atribut "**Hardware**" = $((-3/5 * \log_2 (3/5)) + (-2/5 * \log_2 (2/5)))$
= 0.970950594

- Nilai atribut "**Software**" = $((-1/5 * \log_2 (1/5)) + (-4/5 * \log_2 (4/5)))$
= 0.721928095

- Nilai atribut "**Umum**" = $((-0/4 * \log_2 (0/4)) + (-4/4 * \log_2 (4/4)))$
= 0

c. Entropy Histori Pendidikan

- Nilai atribut “**SMK Komputer**”
 $= ((-2/4 \cdot \log_2 (2/4)) + (-2/4 \cdot \log_2 (2/4))) = 1$
- Nilai atribut “**SMK Teknik**”
 $= ((-0/4 \cdot \log_2 (0/4)) + (-4/4 \cdot \log_2 (4/4))) = 0$
- Nilai atribut “**SMA Umum**”
 $= ((-2/6 \cdot \log_2 (2/6)) + (-4/6 \cdot \log_2 (4/6))) = 0.918295834$

d. Entropy Hobi

- Nilai atribut “**IT**”
 $= ((-4/6 \cdot \log_2 (4/6)) + (-2/6 \cdot \log_2 (2/6))) = 0.918295834$
- Nilai atribut “**Non IT**”
 $= ((-2/8 \cdot \log_2 (2/8)) + (-6/8 \cdot \log_2 (6/8))) = 0.811278124$

e. Entropy Jenis Kelamin

- Nilai atribut “**1**”
 $= ((-4/7 \cdot \log_2 (4/7)) + (-3/7 \cdot \log_2 (3/7))) = 0.985228136$
- Nilai atribut “**0**”
 $= ((-0/7 \cdot \log_2 (0/7)) + (-7/7 \cdot \log_2 (7/7))) = 0$

2. Nilai Gain

Berikut ini adalah nilai Gain dari setiap kriteria.

a. Nilai Gain Minat Calon Mahasiswa

$$= 0.863120569 - ((5/14) \cdot 0.970950594) + ((5/14) \cdot 0.721928095) + ((4/14) \cdot 0) = 0.258521037$$

b. Nilai Gain Histori Pendidikan

$$= 0.863120569 - ((4/14) \cdot 1) + ((4/14) \cdot 0) + ((6/14) \cdot 0.918295834) = 0.183850925$$

c. Nilai Gain Hobi

$$= 0.863120569 - ((6/14) \cdot 0.918295834) + ((8/14) \cdot 0) = 0.005977711$$

d. Nilai Gain Jenis Kelamin

$$= 0.863120569 - ((7/14) \cdot 0.985228136) + ((7/14) \cdot 0) = 0.005977711$$

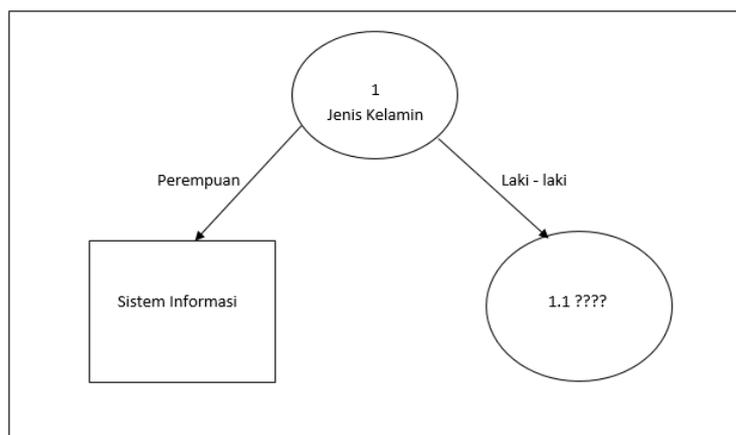
Tabel 2.7. Perhitungan Tahap Pertama

Node		Keterangan	Jml Kasus (S)	Sistem Komputer (SK)	Sistem Informasi (SI)	Entropy	Gain
1	Total		14	4	10	0.863120569	
							0.258521037
	Minat Calon						
		Hardware	5	3	2	0.970950594	
		Software	5	1	4	0.721928095	
		Umum	4	0	4	0	
	Asal Sekolah						0.183850925
		SMK Komputer	4	2	2	1	
		SMK Teknik	4	0	4	0	
		SMA Umum	6	2	4	0.918295834	
	Jenis Kelamin						0.370506501
		Laki - Laki	7	4	3	0.985228136	

Node		Keterangan	Jml Kasus (S)	Sistem Komputer (SK)	Sistem Informasi (SI)	Entropy	Gain
		Perempuan	7	0	7	0	
	Hobi						0.005977711
		IT	6	4	2	0.918295834	
		Non	8	2	6	0.811278124	

(Sumber : Nofriansyah dan Nurcahyo, 2015 : 23)

Tabel 2.5 menunjukkan bahwasanya kriteria jenis kelamin memiliki nilai Gain yang paling tinggi. Untuk fase selanjutnya adalah pembentukan *decision tree*. Berikut ini adalah *decision tree* dari rekapitulasi nilai entropy dan Gain :



Gambar 2.3. Node pertama

(Sumber : Nofriansyah dan Nurcahyo, 2015 : 24)

Pohon keputusan di atas belum terlihat keputusan yang dominan dari setiap program studi yang dipilih. Maka kita harus mencari kembali nilai Entropy dan Gain dari setiap atribut (kriteria) Jenis Kelamin = laki – laki.

1. Nilai Entropy

Berikut ini adalah tabel penyelesaiannya.

Tabel 2.8. Sampel Perhitungan

Kriteria	Atribut	Jumlah Kasus	SI	SK
Jenis Kelamin				
	Laki - laki	7	4	3

(Sumber : Nofriansyah dan Nurcahyo, 2015 : 24)

Setelah itu kita hitung nilai Entropy dari atribut Jenis Kelamin = Laki – laki yang memiliki jumlah kasus “7” seperti terlihat pada tabel 2.5 :

Tabel 2.9. Atribut Jenis Kelamin

No.	Nama Mahasiswa	Minat Calon Mahasiswa	Asal Sekolah	Jenis Kelamin	Hobi	Hasil
1	Novita D.B.B	Hardware	SMK Komputer	Laki - laki	Non	SK
2	Ahmad Riyandi	Hardware	SMK Komputer	Laki - laki	IT	SK
3	Reza Adriansyah	Umum	SMK Komputer	Laki - laki	Non	SI
4	Gafar Dwi Satrio	Software	SMA Umum	Laki - laki	Non	SI
8	Zulfikar Ali	Hardware	SMA Umum	Laki - laki	Non	SK
12	Asri Anzani Br. Tarigan	Umum	SMA Umum	Laki - laki	IT	SI
14	Akbar Widiantara	Software	SMA Umum	Laki - laki	IT	SK

(Sumber : Nofriansyah dan Nurcahyo, 2015 : 24)

Langkah selanjutnya menghitung nilainya, berikut ini adalah rekapitulasi nilai entropy dan gainnya.

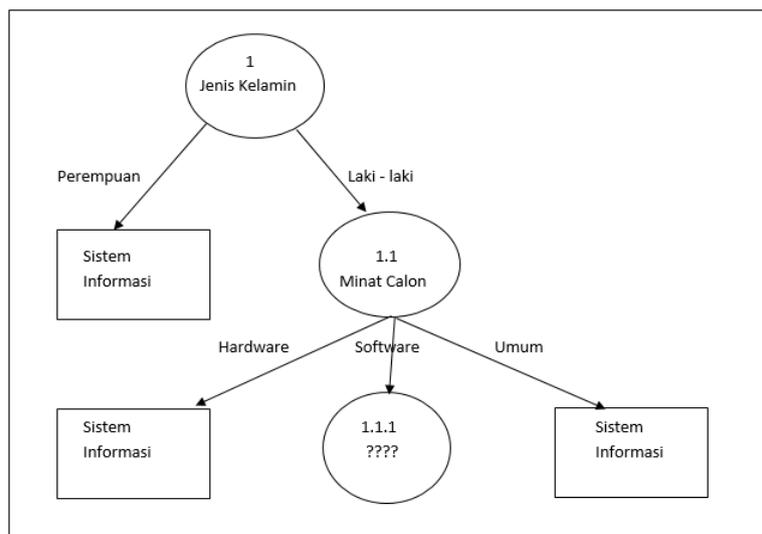
Tabel 2.10. Perhitungan Atribut Jenis Kelamin

Node		Keterangan	Jml Kasus (S)	Sistem Komputer (SK)	Sistem Informasi (SI)	Entropy	Gain
1.1	Jenis Kelamin = Laki - laki		7	4	3	0.98522	
	Minat Calon						0.69951
		Hardware	3	3	0	0	
		Software	2	1	1	1	
		Umum	2	0	2	0	
	Asal Sekolah						0.02024

Node		Keterangan	Jml Kasus (S)	Sistem Komputer (SK)	Sistem Informasi (SI)	Entropy	Gain
		SMK Komputer	3	2	1	0.91829	
		SMK Teknik	0	0	0	0	
		SMA Umum	4	2	2	1	
	Hobi						0.02024
		IT	3	2	1	0.91829	
		Non	4	2	2	1	

(Sumber : Nofriansyah dan Nurcahyo, 2015 : 25)

Berdasarkan tabel diatas terlihat bahwasannya Atribut = Minat Calon memiliki nilai Gain Tertinggi, maka untuk *Root* selanjutnya pada pohon keputusannya dapat terlihat pada gambar (*tree*) berikut ini :



Gambar 2.4. Node Tahap Kedua

(Sumber : Nofriansyah dan Nurcahyo, 2015 : 26)

Karena pohon keputusan belum terlihat keseluruhan hasilnya sehingga kita perlu untuk mencari kembali Nilai Gain dan Entropy selanjutnya berikut ini adalah tabelnya :

Tabel 2.11. Sampel Minat Calon

Kriteria	Atribut	Jumlah Kasus	SK	SI
Minat Calon	Software	2	1	1

(Sumber : Nofriansyah dan Nurcahyo, 2015 : 26)

Setelah itu kita data terlebih dahulu dari atribut Minat Calon = Software yang memiliki jumlah kasus “2” seperti terlihat pada tabel 2.10.

Tabel 2.12. Data Hasil Seleksi

No	Nama Mahasiswa	Minat Calon Mahasiswa	Asal Sekolah	Hobi	Hasil
1	Gadar Swi Satrio	Software	SMA Umum	Non	SI
2	Akbar Widiantara	Software	SMA Umum	TI	SK

(Sumber : Nofriansyah dan Nurcahyo, 2015 : 26)

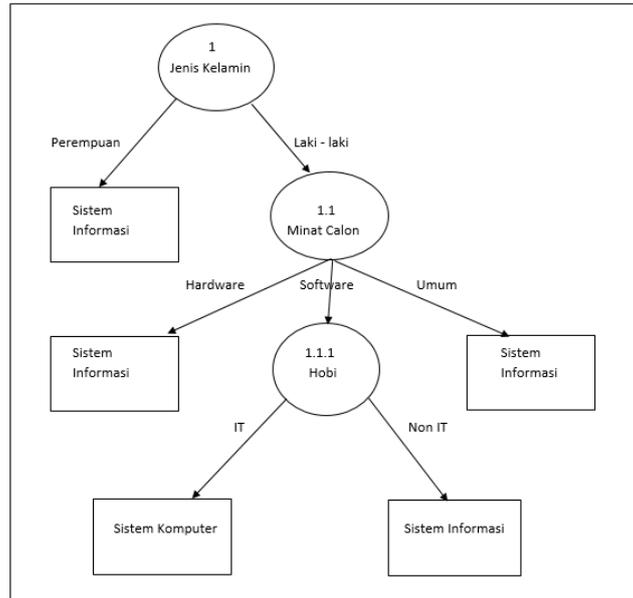
Selanjutnya adalah kita menghitung kembali nilai entropy dan gainnya seperti terlihat pada tabel 2.10. :

Tabel 2.13. Entropy Minat Calon

Node		Keterangan	Jml Kasus (S)	Sistem Komputer (SK)	Sistem Informasi (SI)	Entropy	Gain
1.1.1	Jenis Kelamin = Laki – laki dan Minat Calon = Software		2	1	1	1	
	Asal Sekolah						0
		SMK Komputer	0	0	0	0	
		SMK Teknik	0	0	0	0	
		SMA Umum	2	1	1	1	
	Hobi						1
		IT	1	1	0	0	
		Non IT	1	0	1	0	

(Sumber : Nofriansyah dan Nurcahyo, 2015 : 27)

Pada tabel 2.10. menjelaskan bahwasannya yang memiliki kriteria memiliki nilai Gain Tertinggi yaitu : 1 maka node pohon keputusannya sebagai berikut :



Gambar 2.5. Hasil Pohon Keputusan

(Sumber : Nofriansyah dan Nurcahyo, 2015 : 27)

Maka basis pengetahuan yang terbentuk yaitu :

1. Jika **Jenis Kelamin** = Perempuan maka **Hasil** = Sistem Informasi
2. Jika **Jenis Kelamin** = Laki – laki dan **Minat Calon** = Hardware maka **Hasil** = Sistem Komputer
3. Jika **Jenis Kelamin** = Laki – laki dan **Minat Calon** = Umum maka **Hasil** = Sistem Informasi
4. Jika **Jenis Kelamin** = Laki – laki dan **Minat Calon** = Software dan **Hobi** = IT maka **Hasil** = Sistem Komputer
5. Jika **Jenis Kelamin** = Laki – laki dan **Minat Calon** = Software dan **Hobi** = Non IT maka **Hasil** = Sistem Informasi

D. Tinjauan Pustaka

Penelitian rujukan merupakan acuan yang dibutuhkan seorang peneliti untuk melakukan penelitian. Penelitian rujukan pada penelitian ini diambil berdasarkan kesamaan metode yang digunakan yaitu C4.5 Banyak penelitian yang menggunakan metode ini dalam berbagai kasus. Antara lain :

1. Lareno dan Swastina (2014). Program studi Teknik Informatika, STMIK Banjarmasin Indonesia. **Model Penentuan Potensi Status Gizi Bermasalah Menggunakan *Decision Tree*.**

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan. Dalam penelitian tentang penentuan potensi status gizi bermasalah menggunakan data SKDN 2013, yang didapatkan dari Posyandu Purnama, Desa Pasarlama, Kecamatan Banjarmasin Tengah, Banjarmasin.

Dalam penelitian ini menggunakan algoritma *Decision Tree* merupakan algoritma terbaik. Algoritma *Decision Tree* berbasis data P1, P2, P4 tetap dapat dipakai untuk peta 1, karena telah memiliki akurasi 81,8 %. Peta 2, berbasis P3 dan P5, yang dapat mencapai akurasi 72,7.

2. Hariati, Wati dan Cahyono (2018). Program studi Teknologi Informasi dan Komunikasi, Universitas Mulawarman, Samarinda. **Penerapan Algoritma C4.5 pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kutai Kartanegara.**

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan. Dalam penelitian tentang penentuan penerima program bantuan pemerintah menggunakan data dengan jumlah 123 yang diambil dari data calon penerima bantuan tahun 2014 yang diperoleh dari bidang kesejahteraan sosial Dinas Sosial Kabupaten Kutai Kartanegara.

Dalam penelitian ini menggunakan algoritma *Decision Tree* merupakan algoritma terbaik. Dalam melakukan klasifikasi penentuan penerima program bantuan pemerintah daerah, algoritma C4.5 memiliki nilai akurasi sebesar 97,14% dengan variabel umur, jenis kelamin, status perkawinan, pendidikan terakhir, pekerjaan, dan keterampilan.

3. Irmayansyah dan Firdaus (2018). Program Studi Sistem Informasi, STIKOM BINANIAGA. **Penerapan Algoritma C4.5 Untuk Klasifikasi Penentuan Penerimaan Bantuan Langsung di Desa Ciomas.**

Dalam penelitian ini menggunakan algoritma *Decision Tree* merupakan algoritma terbaik. Data dalam penelitian tentang penentuan penerimaan bantuan langsung adalah data sensus ekonomi penduduk desa ciomas tahun 2015 dan penerima bantuan raskin 2016.

Dalam melakukan klasifikasi penentuan penerima program bantuan langsung, algoritma C4.5 memiliki nilai presentase kelayakan sebesar 74,9% dengan variabel jenis atap rumah, jenis pekerjaan, jenis dinding rumah

terluas, jenis lantai rumah terluas, sumber penerangan utama, sumber air minum, bahan bakar utama untuk memasak, fasilitas tempat buang air besar, status kepemilikan rumah, dan luas rumah.

4. Enstiniaty, Tambunan dan Suhendro (2019). Program Studi Sistem Informasi, STIKOM Tunas Bangsa Pematang Siantar. **Penerapan Data Mining Dalam Menentukan Penerima Bantuan Sosial Dengan Algoritma C4.5 Pada Kantor Kecamatan Siantar Selatan.**

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan. Dalam penelitian tentang penentuan penerima program bantuan sosial menggunakan data masyarakat penerima bantuan sosial dari Kantor Kecamatan Siantar Selatan.

Dalam penelitian ini menggunakan algoritma *Decision Tree* merupakan algoritma terbaik. Pada penelitian ini Algoritma C4.5 berhasil dilakukan dengan menunjukkan bahwa variabel penghasilan menjadi pengaruh tertinggi dalam menentukan bantuan sosial.

5. Agus Bahtiar (2020). Program Studi Teknik Informatika, STMIK IKMI Cirebon. **Penerapan Algoritma *Decision Tree* Untuk Memprediksi Penerima Bantuan Keluarga Harapan.**

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan. Dalam penelitian tentang penentuan penerima program bantuan keluarga harapan menggunakan data penerima bantuan keluarga harapan di kecamatan sumber kabupaten Cirebon dengan jumlah 4347 *record*.

Dalam penelitian ini menggunakan algoritma *Decision Tree* merupakan algoritma terbaik. Pada penelitian ini Algoritma C4.5 memiliki tingkat akurasi sebesar 98,30% dengan variabel nama, balita 0-5 tahun, anak 7-15 tahun, ibu hamil, lansia dan kriteria.

6. Mansyur, Yuyun, dan Alawiyah (2019). Program Studi Sistem Informasi, STMIK Handayani Makassar. **Penerapan Algoritma C4.5 Untuk Klasifikasi Status Kesejahteraan Rumah Tangga Keluarga Binaan Sosial di Kabupaten Bulukumba.**

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan. Dalam penelitian tentang klasifikasi status kesejahteraan keluarga binaan sosial menggunakan data 13.716 Kepala Keluarga, merupakan data sekunder dari penerima bantuan sosial Program Keluarga Harapan (PKH) Kabupaten Bulukumba tahap II tahun 2019.

Dalam penelitian ini menggunakan algoritma *Decision Tree* merupakan algoritma terbaik. Pada penelitian ini Algoritma C4.5 memiliki tingkat akurasi sebesar 89% dengan variabel jenis kelamin, umur, status tempat tinggal, jumlah anggota keluarga, lapangan usaha, jumlah atap terluas, jenis dinding, jenis lantai dan sumber air minum.

7. Hendrian (2018). Program Studi Informatika, Universitas Indraprasta PGRI. **Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan.**

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan. Dalam penelitian tentang prediksi siswa dalam memperoleh bantuan dana pendidikan menggunakan data yang digunakan bersumber dari data Siswa aktif SMA Bina Bangsa Mandiri dengan dengan jumlah kasus sebanyak 254 *record*.

Dalam penelitian ini menggunakan algoritma *Decision Tree* merupakan algoritma terbaik. Pada penelitian ini Algoritma C4.5 memiliki tingkat akurasi sebesar 98, 80% dengan variabel penghasilan ayah, kategori penghasilan ayah, penghasilan ibu, kategori penghasilan ibu.

8. Zaman (2016). Program Studi Komputer dan Teknologi Informasi, Universitas Putra Indonesia "YPTK" Padang. **Penerapan Data Mining Menggunakan Algoritma C4.5 Untuk Menentukan Kelayakan Penerima Bantuan Rehabilitas Sosial Rumah Tidak Layak Huni (Studi Kasus di Pemerintahan Kabupaten Solok Selatan).**

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan. Dalam penelitian tentang penentuan kelayakan penerima bantuan rehabilitas sosial rumah tidak layak huni menggunakan data 30 dengan atribut 7.

Dalam penelitian ini menggunakan algoritma *Decision Tree* merupakan algoritma terbaik. Pada penelitian ini Algoritma C4.5 dapat menemukan pola kelayakan penerima bantuan rehabilitas sosial rumah tidak layak huni dengan variabel status, pekerjaan, pendapatan perbulan, jumlah tanggungan, lantai rumah, bentuk fisik rumah, sarana air bersih dan MCK.

9. Harman (2018). Universitas Putera Batam. **Penerapan Algoritma C4.5 Untuk Kelayakan Penerima Uang dan Beras.**

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan. Dalam penelitian tentang kelayakan penentuan penerima

bantuan menggunakan data calon penerima bantuan beras dan uang diperumahan Taman Hangtuh RW 06 Batam Kota.

Dalam penelitian ini menggunakan algoritma *Decision Tree* merupakan algoritma terbaik. Pada penelitian ini Algoritma C4.5 dapat menemukan perhitungan jumlah kepala keluarga yang menerima bantuan uang dan beras adalah 7 orang atau 43,75% dan yang tidak menerima sebanyak 9 orang atau 56,25% dengan variabel status perkawinan, pekerjaan, jumlah penghasilan, jumlah tanggungan, dan status kepemilikan rumah.

10. Rismayanti (2016). Program Studi Teknik Informatika, Sekolah Tinggi Teknik Harapan Medan. **Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa di STT Harapan Medan.**

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat diambil kesimpulan. Dalam penelitian tentang penentuan penerima beasiswa menggunakan data yang akan digunakan dan dianalisa, yaitu data mahasiswa di Sekolah Tinggi Teknik Harapan (STTH).

Dalam penelitian ini menggunakan algoritma *Decision Tree* merupakan algoritma terbaik. Pada penelitian ini Algoritma C4.5 dapat menghasilkan keputusan mahasiswa yang mendapatkan beasiswa adalah mahasiswa yang bernilai *Cum Laude* dan IPK bernilai *Very Good* dengan tanggungan orang tua bernilai banyak.

Tabel 2.14. Jurnal Referensi

No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
1	Liliana Swastina, Bambang Lareno	Model Penentuan Potensi Status Gizi Bermasalah Menggunakan <i>Decision Tree</i>	Perlunya informasi mengenai wilayah mana saja yang memiliki kurang gizi, sebagai bahan masukan bagi pemerintah dan penentu kebijakan di bidang kesehatan masyarakat dalam melakukan pencegahan dan penanggulangan gizi buruk.	Jurnal GEMA AKTUALITA, Vol. 3 No 2, Desember 2014	Ada kesamaan pada metode yang digunakan adalah metode C4.5 atau <i>decision tree</i> .

No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
2	Hariati, Masna Wati, Bambang Cahyono	Penerapan Algoritma C4.5 Pada Penentuan Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara	Berhubungan dengan banyaknya program bantuan yang ada dengan kriteria yang berbeda-beda dalam penentuan penerimaan program bantuan, tentunya staf pemerintah daerah akan kesulitan dalam melakukan proses seleksi.	JURTI, Vol.2 No.2, Desember 2018, ISSN: 2579-8790	Ada kesamaan pada metode yang digunakan adalah metode C4.5.
3	Aulia Arief Firdaus, Irmayansyah	Penerapan Algoritma C4.5 Untuk Klasifikasi Penentuan Penerimaan Bantuan Langsung di Desa Ciomas	Dalam penyelenggaraan program bantuan pemberian bantuan kepada penduduk tidak mampu masih terdapat masalah pada pelaksanaannya yaitu kurang akuratnya penerima bantuan karena data yang diterima tidak sesuai dengan keadaan penduduk yang ada di berbagai wilayah.	Jurnal Ilmiah Teknologi - Informasi dan Sains (TeknoIS) Volume 8 Nomor 1 Bulan Mei 2018 Hal. 17-28 p-ISSN : 2087-3891 dan e-ISSN : 2597-8918	Ada kesamaan pada metode yang digunakan untuk klasifikasi yaitu algoritma C4.5.
4	Romatua Estiniaty, Heru Satria Tambunan, Dedi Suhendro	Penerapan Data Mining dalam Menentukan Penerima Bantuan Sosial Dengan Algoritma C4.5 Pada Kantor Kecamatan Siantar Selatan	Dalam proses seleksi menentukan penerima bantuan sosial pada kantor Kecamatan Siantar Selatan masih belum sepenuhnya menggunakan teknologi informasi untuk mendukung kinerja pegawai. Sehingga memiliki kendala dan membutuhkan waktu yang cukup lama. Variabel yang digunakan adalah pekerjaan, penghasilan, jumlah tanggungan dan kondisi rumah.	Seminar Nasional Matematika dan Terapan 2019 ISSN: 2721-3684 Volume 1, Desember, pp: 452-460	Ada kesamaan pada metode yang digunakan adalah metode C4.5 atau <i>decision tree</i> .

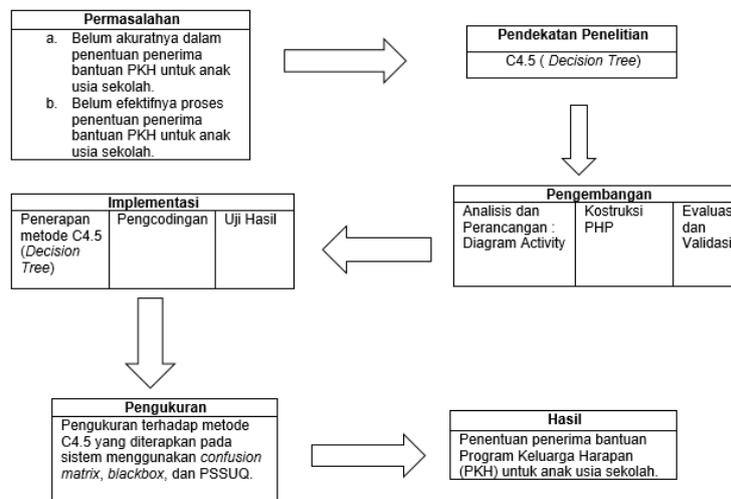
No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
5	Agus Bahtiar	Penerapan Algoritma <i>Decision Tree</i> Untuk Memprediksi Penerima Bantuan Keluarga Harapan	Permasalahan sering terjadi terhadap program bantuan pemberdayaan dari pemerintah salah satunya pada bantuan PKH masih saja terjadi tidak tepat sasaran warga yang menerima bantuan tersebut. Variabel yang digunakan pada penelitian ini adalah balita 0-5 tahun, anak usia 7-15 tahun, lansia >60 tahun, ibu hamil.	Jurnal ICT : Information Communication & Technology Vol. 19, No.1, Juli 2020, pp.70-76 p-ISSN: 2302-0261, e-ISSN: 2303-3363, DOI: 10.36054/jict-ikmi.v19i1.93	Ada kesamaan pada variabel yang digunakan. Kesamaan variabel adalah variabel anak usia 7-15 tahun atau anak usia sekolah.
6	Masnyur, Yuyun, Rabiatul Adawiyah	Penerapan Algoritma C4.5 Untuk Klasifikasi Kesejahteraan Rumah Tangga Keluarga Binaan Sosial di Kabupaten Bulukumba	Kemiskinan merupakan salah satu permasalahan dalam upaya peningkatan kesejahteraan di semua daerah. Variabel yang digunakan pada penelitian ini adalah jenis kelamin, umur, status tempat tinggal, jumlah anggota keluarga, lapangan usaha, jumlah atap terluas, jenis dinding, jenis lantai, sumber air minum.	Jurnal Teknologi Informasi dan Komunikasi P-ISSN : 2088-6705 Volume 9, Nomor 2, Desember 2019 : 147 – 154 E-ISSN : 2621-5608	Ada kesamaan pada metode yang digunakan adalah metode C4.5 atau <i>decision tree</i> .
7	Senna Hendrian	Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan	Pendidikan merupakan salah satu komponen kehidupan yang dapat menunjang keberhasilan seseorang menuju kehidupan yang jauh lebih baik lagi. Terutama bagi anak yang ada dalam lingkup usia wajib belajar. akan tetapi tidak semua anak wajib belajar dapat mengikuti pendidikan, karena beberapa faktor penyebab, salah satunya adalah masalah biaya pendidikan.	Faktor Exacta 11 (3): 266-274, 2018 p-ISSN: 1979-276X e-ISSN: 2502-339X DOI : 10.30998/faktorexacta.v11i3.2777	Ada kesamaan pada metode yang digunakan adalah metode C4.5 atau <i>decision tree</i> .

No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
8	Khairil Zaman	Penerapan Data Mining Menggunakan Algoritma C4.5 Untuk Menentukan Kelayakan Penerima Bantuan Rehabilitas Sosial Rumah Tidak Layak Huni (Studi Kasus di Pemerintahan Kabupaten Solok Selatan)	Untuk menentukan kelayakan penerima bantuan RSRTLH pada Dinas Sosial Ketenagakerjaan dan Transmigrasi Kabupaten Solok Selatan. data dikumpulkan melalui observasi dan interview yang dilakukan kepada instansi yang bersangkutan. Variabel yang digunakan status perkawinan, pekerjaan, pendapatan /bulan, jumlah tanggungan, lantai rumah, bentuk fisik rumah, sarana air bersih dan MCK.	UPI YPTK Jurnal KomTekInfo Vol. 3, No. 2, Desember 2016, Hal. 12-24 ISSN :2356-0010 Copyright©2016 by LPPM UPI YPTK Padang	Ada kesamaan pada metode yang digunakan adalah metode C4.5 atau <i>decision tree</i> .
9	Rika Harman	Penerapan Algoritma C4.5 Untuk Kelayakan Penerimaan Uang dan Beras	Pemberian uang dan beras bagi warga yang layak menerima. Banyak warga yang tidak memasuki kategori dan menerima bantuan berupa uang dan beras. Variabel yang ada di jurnal ini adalah status perkawinan, pekerjaan, jumlah penghasilan, jumlah tanggungan, status kepemilikan rumah.	CBIS JOURNAL - VOL. 06 NO. 01 (2018) : MARET	Ada kesamaan pada metode yang digunakan adalah metode C4.5 atau <i>decision tree</i> .
10	Rismayanti	Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa di STT Harapan Medan	Pemberian bantuan belajar berupa beasiswa juga diberikan kepada mahasiswa di STT-Harapan Medan, salah satu beasiswa yang diberikan adalah Beasiswa Peningkatan Prestasi Akademik (PPA) yang diberikan kepada mahasiswa berprestasi, di mana IPK menjadi kriteria utama dalam pemberian beasiswa tersebut.	Jurnal Media Infotama Vol. 12 No. 2, September 2016	Ada kesamaan pada metode yang digunakan adalah metode C4.5 atau <i>decision tree</i> .

Berdasarkan uraian maka terdapat kontribusi dari penelitian adanya kesamaan variabel dan kesamaan metode yang digunakan yaitu klasifikasi metode algoritma C4.5 atau metode *decision tree*. Kontribusi lainnya adalah bidang yang digunakan adalah bidang yang sama yaitu bantuan sosial.

E. Kerangka Pemikiran

Berikut adalah kerangka pemikiran untuk memecahkan masalah penelitian ini yang digambarkan pada gambar berikut:



Gambar 2.6. Kerangka Pemikiran

Kerangka pemikiran pada gambar diatas dapat dijelaskan sebagai berikut :

- Masalah dari penelitian ini dapat dikenali, permasalahannya adalah belum akuratnya dalam penentuan penerima bantuan PKH untuk anak usia sekolah serta belum efektifnya proses penentuan bantuan PKH untuk anak usia sekolah.
- Pendekatan metode yang digunakan adalah metode C4.5 yang berfungsi untuk mengklasifikasian penerima bantuan Program Keluarga Harapan (PKH) untuk anak usia sekolah.
- Pengembangan dilakukan dengan tiga tahap, yaitu analisis dan perancangan diagram activity, kontruksi PHP, dan dilakukan evaluasi dan validasi.
- Setelah itu dilakukan implementasi yang dilakukan penerapan metode, pengcodingan, dan uji hasil.
- Pengukuran digunakan menggunakan *confussion matrix*, blackbox dan PSSUQ.
- Proses yang dilakukan akan menghasilkan penerima bantuan Program Keluarga Harapan (PKH) untuk anak usia sekolah.

F. Hipotesis Penelitian

Berdasarkan permasalahan yang dihadapi yaitu menentukan penerima bantuan Program Keluarga Harapan (PKH) untuk anak usia sekolah yang memerlukan bantuan, belum efektif dan belum akurat pada saat proses klasifikasi bantuan Program Keluarga Harapan (PKH) untuk anak usia sekolah, maka perlu adanya suatu cara untuk mengatasi permasalahan tersebut. Dalam teori data mining ada beberapa metode yang dapat melakukan pengklasifikasian untuk memprediksi suatu peluang di masa depan berdasarkan dari pengalaman di masa sebelumnya, diantaranya adalah metode C4.5. Teori klasifikasi yang digunakan adalah C4.5 yang akan menghasilkan hasil prediksi yang telah ditentukan nilai probabilitas akhir dari setiap kelas atau label dan melihat hasil akhir yang terbesar dari setiap kelas nantinya hasil penentuan tersebut bisa digunakan untuk menentukan penerima bantuan anak usia sekolah. Berdasarkan dari pengungkapan penulis, maka dapat ditetapkan hipotesis metode C4.5 diduga dapat klasifikasi penerima bantuan Program Keluarga Harapan (PKH) untuk anak usia sekolah.