

## **BAB II Kerangka Teoritis**

### **A. Landasan Teori**

#### **1. Pengertian Data Mining**

Menurut (Turban dkk, 2001), Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran computer (machine learning) untuk menganalisis dan mengekstrasi pengetahuan (knowledge) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (induction-based learning) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep - konsep yang akan dipelajari. Knowledge Discovery in Databases (KDD) adalah penerapan metode saintifik pada data mining. Dalam konteks ini data mining merupakan satu langkah dari proses KDD.

#### **2. Pengertian Klasifikasi**

Menurut (Prasetyo, 2013), Klasifikasi dapat didefinisikan secara detail sebagai waktu pekerjaan yang melakukan pelatihan atau pembelajaran terhadap fungsi target / yang memetakan setiap vektor (set fitur)  $x$  ke depan satu dari sejumlah tabel kelas  $y$  yang tersedia pekerjaan pelatihan tersebut akan menghasilkan suatu model yang kemudian disimpan sebagai memori. Klasifikasi bertujuan untuk mengklasifikasikan item data menjadi satu dari beberapa kelas standar. Sebagai contoh, suatu program email dapat mengklasifikasikan email yang sah dengan email spam. Beberapa algoritma klasifikasi antara lain pohon keputusan, nearest neighbor, naive bayes, neural networks dan support vector machines.

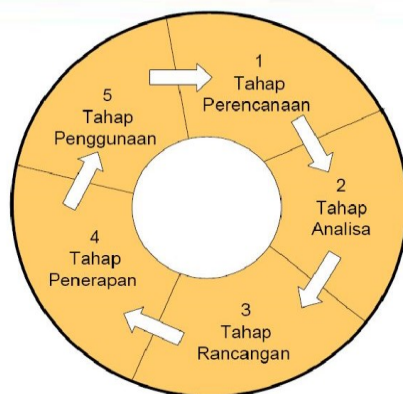
#### **3. SDLC**

Menurut (Firmansyah, 2018) mengemukakan bahwa SDLC atau Software Development Life Cycle atau sering disebut juga System Development Life Cycle adalah proses mengembangkan atau mengubah suatu sistem perangkat lunak dengan menggunakan model-model dan metodologi yang digunakan orang untuk mengembangkan sistem-sistem perangkat lunak sebelumnya, berdasarkan best practice atau cara-cara yang sudah teruji baik. Sedangkan menurut Sukamoto dan Shalahuddin di jelaskan bahwa model waterfall sering juga disebut model sekuensi linear atau alur hidup klasik. Pengembangan sistem dikerjakan secara terurut mulai dari analisis, desain,

pengkodean, pengujian dan tahap pendukung. Terdapat beberapa tahapan pekerjaan pengembangan yang perlu dilakukan jika suatu proyek ingin memiliki kemungkinan berhasil yang besar. Tahapan-tahapan tersebut adalah:

- 1) Perencanaan
- 2) Analisis
- 3) Desain
- 4) Implementasi
- 5) Penggunaan

Proyek dan sumber daya yang dibutuhkan untuk melakukan pekerjaan direncanakan kemudian disatukan. Sistem yang ada juga dianalisis untuk memahami masalah dan menentukan persyaratan fungsional dari sistem yang baru. Sistem baru ini kemudian dirancang dan diimplementasikan. Setelah implementasi, sistem kemudian digunakan, idealnya untuk jangka waktu yang lama, karena pekerjaan-pekerjaan di atas mengikuti satu pola yang teratur dan dilaksanakan dengan cara dari atas ke bawah, SDLC tradisional sering kali disebut sebagai pendekatan air terjun (*waterfall approach*). Aktivitas ini memiliki aliran satu arah menuju ke penyelesaian proyek.



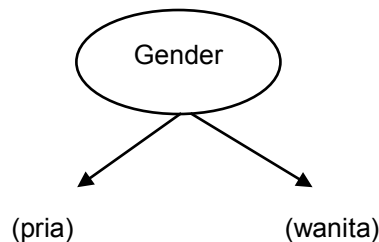
Gambar 2. 1 Pola lingkaran  
(Sumber: Raymond McLeod 2007: 199)

Metode SDLC selanjutnya adalah prototype. Metode prototype adalah metode yang memungkinkan pengguna atau *user* memiliki gambaran awal tentang perangkat lunak yang akan dikembangkan, serta pengguna dapat melakukan pengujian di awal sebelum perangkat lunak dirilis. Metode ini bertujuan untuk mengembangkan model menjadi perangkat lunak yang final. Artinya sistem akan dikembangkan lebih cepat dan biaya yang dikeluarkan

lebih rendah. Metode prototype ini memiliki tahap-tahap yang harus dilakukan dalam pengembangan perangkat lunak.

#### 4. Metode Algoritma C 4.5

Menurut (Novita, 2016), Algoritma yang dapat digunakan untuk memprediksi atau mengklasifikasi suatu kejadian dengan pembentukan pohon keputusan antara lain algoritma C4.5, yang merupakan salah satu algoritma induksi pohon keputusan yang dikembangkan oleh J.Ross Quinlan. Algoritma C4.5 merupakan kelompok algoritma decision tree. Algoritma ini mempunyai input berupa training samples dan samples. Training samples merupakan data contoh yang digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang digunakan sebagai parameter dalam klasifikasi data.



Gambar 2. 2 syarat pengujian fitur biner

Yang menjadi hal penting dalam induksi decision tree adalah bagaimana menyatakan syarat pengujian pada node. Ada 3 kelompok penting dalam syarat pengujian node:

a) Fitur biner

Fitur yang hanya mempunyai dua nilai berbeda disebut dengan fitur biner. Syarat pengujian ketika fitur ini menjadi node (akar maupun internal) hanya punya dua pilihan cabang. Contoh pemecahannya disajikan pada Gambar 2.2.

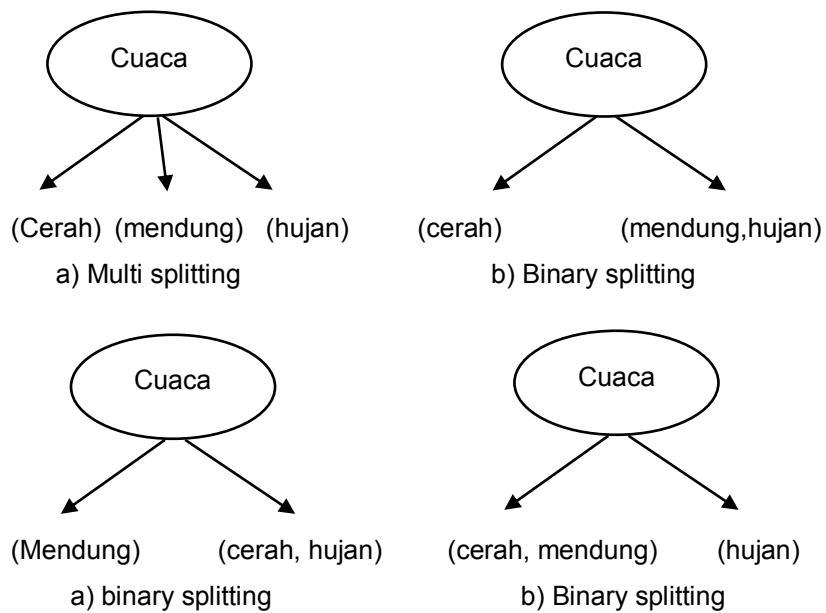
b) Fitur bertipe kategorikal

Untuk fitur yang nilainya bertipe kategorikal (nominal atau ordinal) bisa mempunyai beberapa nilai berbeda. Contohnya adalah fitur 'cuaca' mempunyai 3 nilai berbeda dan ini bisa mempunyai banyak kombinasi syarat pengujian pemecahan. Secara umum ada 2, yaitu pemecahan biner (binary splitting) dan (multi splitting). Kombinasinya disajikan seperti pada Gambar 2.3. Untuk pemecahan yang hanya membolehkan pemecahan biner, seperti algoritma CART, maka akan memberikan

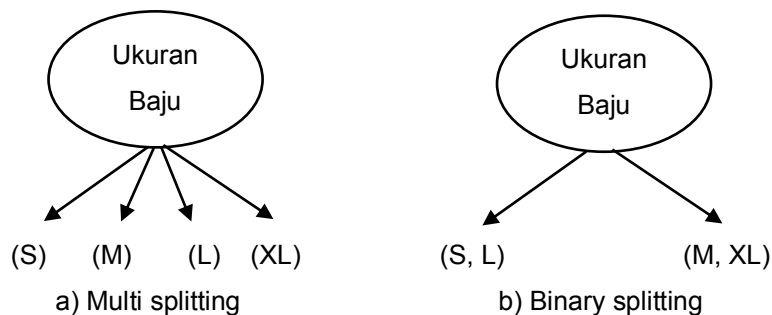
kemungkinan jumlah kombinasi pemecahan sebanyak  $2^{(k-1)}$ , dimana k adalah jumlah nilai berbeda dalam fitur tersebut. Contoh pemecahannya disajikan pada Gambar 2.3 dan Gambar 2.4.

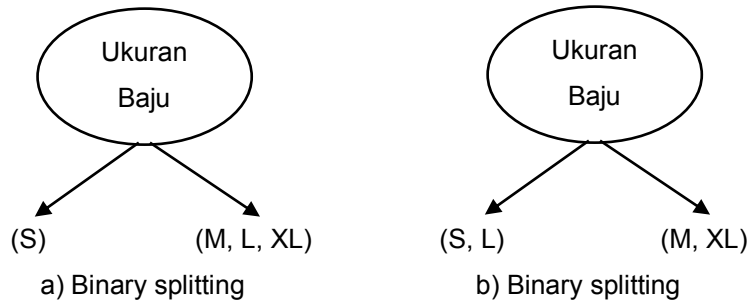
c) Fitur bertipe numerik

Untuk fitur bertipe numerik, syarat pengujian dalam node (akar maupun internal) dinyatakan dengan pengujian perbandingan ( $A < v$ ) atau ( $A \geq v$ ) dengan hasil biner, atau untuk multi dengan hasil berupa jangkauan nilai dalam bentuk  $v_i \leq A < v_{(i+1)}$ , untuk  $i = 1, 2, \dots, k$ . Untuk kasus pemecahan biner, maka algoritma akan memeriksa semua kemungkinan posisi pemecahan  $v$  dan memilih posisi  $v$  terbaik. Untuk cara multi, maka algoritma harus memeriksa semua kemungkinan jangkauan nilai kontinyu. Contoh pemecahan pada fitur numerik disajikan pada Gambar 2.5.

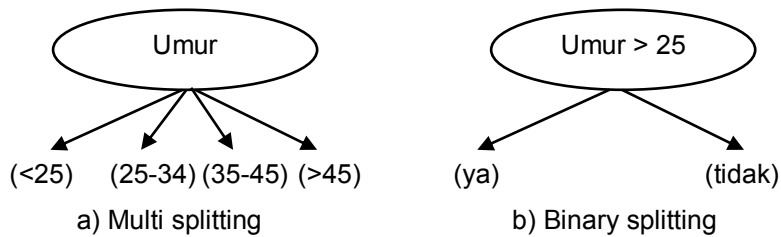


Gambar 2. 3 Syarat pengujian fitur bertipe nominal





Gambar 2. 4 Syarat pengujian fitur bertipe ordinal



Gambar 2. 5 Syarat pengujian fitur bertipe numerik

Kriteria yang paling banyak digunakan untuk memilih fitur sebagai pemecah dalam algoritma C 4.5 adalah rasio gain, yang diformulasikan oleh persamaan berikut:

$$RasioGain(s, j) = \frac{Gain(s, j)}{SplitInfo(s, j)}$$

Menurut (Prasetyo, 2014), persamaan diatas menyatakan nilai rasio gain pada fitur ke- $j$ .  $SplitInfo(s, j)$  dapat dari  $SplitInfo(s, j) = -\sum_{i=1}^k p(v_i|s) \log_2 p(v_i|s)$  dimana  $k$  menyatakan jumlah pemecahan.



## 5. PHP




Menurut (Anhar, 2010) PHP ialah bahasa pemrograman web server- side yang bersifat open source, dan bahasa pemrograman yang open source sangat cocok atau dikhususkan dalam pengembangan web dan dapat ditanamkan pada sebuah laporan penelitian html PHP atau singkatnya Hypertext Preprocessor adalah bahasa pemrograman open source yang sangat atau sangat cocok untuk pengembangan web dan dapat disematkan dalam makalah HTML. PHP adalah bahasa skrip sisi server yang memproses data di sisi server. Pengertian lain dari PHP adalah singkatan dari "Hypertext Preprocessor".

## 6. Unified Modeling Language (UML)

Menurut (Rosa & Shalahuddin, 2016) Unified Modeling Language (UML) adalah bahasa pemodelan yang digunakan untuk mendefinisikan atau menggambarkan sistem perangkat lunak berdasarkan objek dalam sistem. UML tidak menentukan metode yang akan digunakan saat mengembangkan sistem, hanya notasi standar yang biasa digunakan untuk pemodelan objek. Use Case Diagram adalah pemodelan untuk kelakuan (behavior) sistem informasi yang akan dibuat. Use Case mendeskripsikan sebuah interaksi antara satu atau lebih actor dengan sistem informasi yang akan dibuat. Berikut adalah simbol dari use case diagram.







Tabel 2. 1 Simbol use case diagram

SIMBOL	NAMA	KETERANGAN
	<i>ACTOR</i>	Orang proses atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat diluar sistem informasi dibuat itu sendiri, jadi walaupun simbol dari actor adalah gambar orang, biasanya dinyatakan menggunakan kata benda di awal frase nama actor.
	<i>USE CASE</i>	Fungsionalitas yang disediakan sistem sebagai unit-unit yang saling bertukar pesan antar unit atau actor biasanya dinyatakan dengan menggunakan kata kerja di awal frase nama use case.



<b>SIMBOL</b>	<b>NAMA</b>	<b>KETERANGAN</b>
	<p><i>ASOSIASI//</i> <i>ASSOCIATION</i></p>	Komunikasi antara actor dan use case yang berpartisipasi pada use case atau use case memiliki interaksi dengan actor.
<p>&lt;&lt;extend&gt;&gt;  </p>	<p><i>EKSTENSI//</i> <i>EXTEND</i></p>	Relasi use case tambahan ke sebuah use case dimana use case yang ditambahkan dapat berdiri sendiri walau tanpa use case tambahan memiliki nama depan yang sama dengan use case yang ditambahkan.
	<p><i>GENERALISASI //</i> <i>GENERALIZATION</i></p>	Hubungan generalisasi dan spesialisasi (umum-khusus) antara dua buah use case dimana fungsi yang satu adalah fungsi yang lebih umum dari lainnya.
<p>&lt;&lt;include&gt;&gt;  </p>	<p><i>MENGGUNAKAN</i> <i>INCLUDE</i></p>	Relasi usecase tambahan ke sebuah use case untuk menjalankan fungsional.

Tabel 2. 2 Simbol Activity Diagram

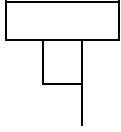







<b>SIMBOL</b>	<b>NAMA</b>	<b>KETERANGAN</b>
---------------	-------------	-------------------

<b>SIMBOL</b>	<b>NAMA</b>	<b>KETERANGAN</b>
	<i>STATUS AWAL/INITIAL</i>	Status awal aktivitas sistem, sebuah diagram aktivitas memiliki sebuah status awal.
	<i>AKTIVITAS/ ACTIVITY</i>	Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja.
	<i>PERCABANGAN/ DECISION</i>	Asosiasi percabangan dimana lebih dari satu aktivitas digabungkan menjadi satu.
	<i>PENGGABUNGAN/ JOIN</i>	Asosiasi penggabungan dimana lebih dari satu aktivitas lebih dari satu.
	<i>STATUS AKHIR/FINAL</i>	Status akhir yang dilakukan sistem, sebuah diagram aktivitas memiliki sebuah status baru.
	<i>SWIMLINE</i>	Memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas yang terjadi.

Tabel 2. 3 Simbol Sequence Diagram

<b>SIMBOL</b>	<b>NAMA</b>	<b>KETERANGAN</b>
	<i>Actor</i>	Merepresentasikan entitas yang berada diluar sistem dan berinteraksi diluar sistem.
	<i>Lifeline</i>	Menghubungkan objek selama sequence (message dikirim atau diterima).
		Merepresentasikan entitas



SIMBOL	NAMA	KETERANGAN
	<i>General</i>	tunggal dalam sequence.
	<i>Boundary</i>	Berupa tepi dari sistem, seperti user interface dan alat yang berinteraksi dengan yang lain.
	<i>Control</i>	Elemen mengatur aliran dari informasi untuk sebuah skenario. Objek ini umumnya perilaku dan perilaku bisnis.
	<i>Entitas</i>	Elemen yang bertanggung jawab menyimpan atau informasi. Ini dapat berupa beans atau model object.
	<i>Activation</i>	Suatu titik dimana sebuah objek mulai berpartisipasi dalam sebuah sequence yang menunjukkan sebuah objek mengirim atau menerima objek.
	<i>Message Entry</i>	Berfungsi untuk menggambarkan pesan /hubungan antar objek yang menunjukkan urutan kejadian
	<i>Message to Self</i>	Simbol ini menggambarkan pesan/hubungan objek itu sendiri, yang menunjukkan urutan kejadian yang terjadi
	<i>Message Return</i>	Menggambarkan hasil dari pengiriman message yang digambarkan dengan arah dari kanan ke kiri.

## 7. Pengertian Pembayaran SPP

Menurut (Sarmidi & Fahmi, 2019), Sistem pembayaran dapat diartikan sebagai salah satu kegiatan atau usaha untuk membantu, melayani, mengarahkan atau mengatur semua kegiatan di dalam mencapai suatu tujuan. Karena pembayaran merupakan unsur penting bagi sebuah instansi maka di perlakukan suatu sistem yang dapat mengelola sistem pembayaran secara baik. Sumbangan Pembinaan Pendidikan(SPP) merupakan iuran rutin sekolah yang mana pembayarannya dilakukan setiapsebulan sekali. SPP merupakan salah satu bentuk kewajiban setiap siswa yang masih aktif disekolah tersebut.

## B. Tinjauan Studi

Penelitian rujukan merupakan acuan yang dibutuhkan seorang peneliti untuk melakukan penelitian. Penelitian rujukan pada penelitian ini diambil berdasarkan kesamaan metode yang digunakan yaitu metode algoritma C 4.5. Banyak penelitian yang menggunakan metode ini dalam berbagai kasus. Antara lain :

1. Menurut Hendrian,(2018) penelitiannya yang berjudul **“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. Untuk mengatasi permasalahan yang ada, maka sekolah Bina Bangsa Mandiri menyusun program Bantuan Dana Pendidikan bagi Siswa yang dianggap kurang mampu secara strata ekonomi. Pada penelitian ini penulis menggunakan Algoritma Klasifikasi Datamining yaitu Algoritma C4.5 untuk memprediksi siswa dalam memperoleh bantuan dana pendidikan. Sampel data diambil dari SMA Bina Bangsa Mandiri yang beralamat di Kecamatan Gunungputri Kab. Bogor. Dari hasil pengujian digunakan tes Cros Validation dan Confusion Matrix dan Kurva ROC. Hasil yang diperoleh untuk nilai Accuracy Algoritma C4.5 adalah sebesar 98,80%, nilai untuk Precision sebesar 98,02%, dan nilai untuk Sensitivity atau Recall sebesar 99,00%. Dengan demikian Algoritma C4.5 merupakan algoritma dan teknik terbaik untuk Memprediksi Siswa dalam memperoleh Bantuan Dana Pendidikan.
2. Menurut Maurina(2016) penelitiannya yang berjudul **“PENERAPAN DATA MINING UNTUK REKOMENDASI BEASISWA PADA SMA MUHAMMADIYAH**

**GUBUG MENGGUNAKAN ALGORITMA C4.5**". Klasifikasi rekomendasi beasiswa dilakukan untuk mengklasifikasi apakah siswa akan mendapatkan rekomendasi beasiswa sesuai dengan bobot yang akan di nilai. Klasifikasi dilakukan menggunakan data mining algoritma C4.5. Data yang digunakan yaitu data jurusan, kelas, jumlah nilai, penghasilan orangtua, dan jumlah saudara kandung. Proses data mining pada data training akan menghasilkan pohon keputusan atau rule. Metode evaluasi yang dilakukan dalam penelitian ini yaitu menggunakan confusion matrix dan nilai akurasi, untuk sekali pengujian tingkat akurasi yang dihasilkan yaitu 77%. Hal ini membuktikan bahwa algoritma C4.5 cukup akurat dalam menentukan rekomendasi beasiswa pada SMA Muhammadiyah Gubug.

3. Menurut Isyriyah, Sakaria(2016) penelitiannya yang berjudul "**PERFORMANCE METODE DECISION TREE C 4.5 PADA ANALISA PENGAMBILAN KEPUTUSAN PENERIMA BEASISWA**" Beasiswa adalah program yang dimiliki oleh hampir semua instansi pendidikan, baik SMA/SMK, namun dengan banyaknya kriteria dalam pemilihan penerima beasiswa, dibandingkan dengan jumlah siswa yang banyak serta keadaan mereka yang beragam, menjadikan proses pemilihan penerima beasiswa berjalan lambat dan sering kali hasil seleksinya kurang tepat sasaran. Dari gagasan masalah yang telah digambarkan diatas, dapat disimpulkan bahwa dibutuhkan suatu program yang bertujuan memperoleh hasil penerima beasiswa yang tepat sasaran dalam waktu yang lebih cepat, yaitu dengan sistem penunjang keputusan berbasis web dengan metode Decision Tree C4.5, diharapkan dapat membantu dalam menangani masalah yang ada sekarang. Metode Decision Tree C4.5 merupakan suatu metode pengambilan keputusan untuk menyelesaikan masalah pemilihan beasiswa. Dimana dalam program ini terdapat data training yang digunakan untuk membuat aturan dalam pemilihan beasiswa. Algoritma klasifikasi data mining dengan model algoritma C4.5 dilakukan dengan pengujian-pengujian yang terukur melalui uji AUC, ROC dan T-Test dengan bantuan rapid miner. Hasilnya, setelah dilakukan pengujian dengan menggunakan parameter mahasiswa penerima mahasiswa dengan jumlah mahasiswa sebagai sampel 150 ternyata menghasilkan akurasi sebesar secara keseluruhan nilai hasil validasi adalah accuracy = 90,06%, precision = 100,00% dan recall = 75,00%, artinya akurasi pengujian dengan menggunakan algoritma C4.5 masih baik dan dapat dijadikan salah satu pedoman untuk seleksi penerima beasiswa sebelum ada keputusan penerima beasiswa. Hasil pengujian dengan menggunakan algoritma C4.5 bukan satu-satu algoritma yang dapat digunakan untuk

seleksi penerima beasiswa, masih banyak algoritma lain yang dapat diperbandingkan dengan algoritma C4.5 dimana kemungkinan hasilnya akan berbeda.

4. Menurut Badriah, Nugroho, Sanjaya, Rismawati, Rozikin(2020) penelitiannya yang berjudul **“KLASIFIKASI ALGORITMA C4. 5 DALAM MENENTUKAN PENERIMA BANTUAN COVID- 19(Studi Kasus : Desa di Karawang)”** Covid-19 menyebabkan masyarakat Indonesia mengalami penurunan dalam pendapatan penghasilan. Dalam menanggulangi masalah penurunan pendapatan tersebut pemerintah memberikan bantuan bagi masyarakat akibat terdampak Covid-19. Berhubungan dengan adanya bantuan Covid-19 tersebut, staf pemerintah desa kesulitan dalam melakukan proses seleksi penerima bantuan dan adanya ketidaktepatan dalam penerima bantuan Covid-19. Tujuan penelitian ini adalah membuat klasifikasi penerima bantuan sosial Covid-19 sebagai bahan acuan dalam kebijakan penentuan penerima bantuan sosial. Sehingga dilakukan penelitian untuk mempermudah dalam penentuan dan ketepatan dalam bantuan Covid-19. Algoritma klasifikasi yang digunakan yaitu algoritma C4.5 dengan bahasa pemrograman php. Data yang digunakan terdapat 430 data dengan 5 atribut yaitu jumlah anggota keluarga, pekerjaan, penghasilan sebelum Covid-19, penghasilan setelah Covid-19 dan keputusan (data per bulan April 2020). Hasil dari penelitian ini didapat rule dari algoritma C4.5 dengan dengan nilai akurasi yang terbaik terdapat pada perbandingan 90% data training dan 10% data testing dengan nilai akurasi sebesar 79,54%.
5. Menurut Amaliyanisa(2020) penelitiannya yang berjudul **“PEMILIHAN CALON PENERIMA BANTUAN SISWA MISKIN MENGGUNAKAN ALGORITMA C4.5 DI SMAN 1 AEK KUASAN”**. Program Bantuan Siswa Miskin (BSM) merupakan Program Nasional yang bertujuan untuk mengurangi halangan siswa miskin dalam berpartisipasi untuk bersekolah dengan memberikan bantuan berupa dana dan mendapatkan pendidikan yang layak, mencegah siswa putus sekolah, membantu siswa memenuhi kebutuhan dalam kegiatan pembelajaran, mendukung program wajib belajar 9 tahun (bahkan hingga tingkat menengah atas) serta membantu kelancaran program sekolah. Untuk mendapatkan dana Bantuan Siswa Miskin (BSM) tersebut, pemerintah menetapkan beberapa kriteria siapa saja siswa yang berhak mendapatkan bantuan tersebut. Cara mengatasi untuk melakukan penyeleksian dari Kriteria yang sudah ditetapkan yaitu dengan memanfaatkan teknik Data Mining. Data mining juga merupakan salah satu bagian proses Knowledge Discovery from Databases (KDD) yang bertugas untuk mengekstrak pola atau model dari data dengan menggunakan

suatu algoritma yang spesifik didalam penerapan data mining ini dibantun dengan metode Algroitma C4.5. Algoritma C4.5 membentuk pohon keputusan yang nantinya dapat digunakan sebagai dasar pertimbangan dalam menentukan penerima Bantuan Siswa Miskin (BSM). Tujuan penelitian ini dilakukan untuk membantu kepala sekolah SMAN 1 Aek Kuasan dalam penyeleksian pemilihan calon siswa penerima BSM (Bantuan Siswa Miskin). Adapun batasan dalam penelitian ini data yang diimplementasikan dalam membuat aplikasi ini adalah Visual Basic 2010. Metode penelitian menggunakan metode kualitatif, dimana sebuah penelitian riset yang sifatnya deskripsi, mengacu pada data, memanfaatkan teori yang ada sebagai bahan pendukung, serta menghasilkan suatu teori. Dari hasil penelitian ini sendiri nantinya akan dibuat sistem yang dapat membantu pihak sekolah agar lebih mempermudah dalam penyeleksian pemilihan calon siswa penerima Bantuan Siswa Miskin (BSM).

6. Menurut Merdekawati, Rahayu(2020) penelitiannya yang berjudul **“SISTEM PENDUKUNG KEPUTUSAN PEMBERIAN BANTUAN EKONOMI MENGGUNAKAN ALGORITMA C 4.5”**. Permasalahan yang sering timbul dalam pemberian bantuan ekonomi adalah distribusi bantuan yang tidak tepat sasaran. Problem tidak tepat sasaran ini dialami oleh Baitul Maal Yayasan Al-Anshor Asmawiah sebuah yayasan yang menerima dan menyalurkan bantuan ekonomi. Kendala yang mereka hadapi dikarenakan banyaknya atribut kriteria calon penerima sehingga membingungkan pihak yayasan dalam melakukan seleksi bantuan dan beberapa pihak masih berbuat curang. Informasi yang diberikan calon penerima bantuan sering tidak lengkap sehingga terjadi kekurangan data pada form bantuan ekonomi. Metode untuk menyeleksi bantuan ekonomi ini yaitu Algoritma C4.5 dimana algoritma ini merupakan algoritma yang dapat membentuk decision tree (pohon keputusan) yang digunakan sebagai acuan seleksi bantuan ekonomi. Setelah mendapatkan rules, selanjutnya diimplementasikan dalam sebuah sistem agar memudahkan pengguna memberikan bantuan. Sistem ini mempermudah masyarakat dalam mengisi form bantuan ekonomi dan mudah langsung mengetahui informasi apakah berhak mendapatkan bantuan ekonomi atau tidak serta memudahkan administrator mendapatkan informasi yang berhak menerima atau tidak menerima bantuan.
7. Menurut Azahari, Nursobah(2021) penelitiannya yang berjudul **“Rekomendasi Penerimaan Beasiswa Yayasan Untuk Siswa Baru SMK TI Airlangga dengan Algoritma C4.5”**. Jalur beasiswa yayasan merupakan fasilitas yang disediakan pihak swasta untuk membantu program pendidikan pemerintahan mewujudkan wajib belajar 12 tahun yang salah satunya adalah fasilitas tiga

tahun jenjang pendidikan sekolah menengah kejuruan dan pastinya yayasan akan memperhitungkan siapa yang pantas untuk mendapatkan kesempatan belajar selama tiga tahun untuk siswa baru hal ini akan membantu banyak pihak yang benar-benar pantas mendapatkan kesempatan belajar untuk menumbuhkan semangat belajarnya kembali, tetapi dalam hal ini juga pihak yayasan tentu harus melakukan seleksi terhadap kandidat yang mengajukan diri mendapatkan beasiswa tersebut tentu banyak yang ingin mendapatkan beasiswa tersebut sehingga benar-benar harus dilakukan pertimbangan yang akurat agar beasiswa tidak jatuh kepada tangan yang salah, salah satu cara yang tepat adalah melakukan klasifikasi terhadap data-data yang masuk dengan menggunakan penerapan algoritma C4.5 sehingga proses seleksi nantinya tidak memerlukan waktu yang sangat lama dan berdasarkan penilaian, hasil akhir dari algoritma C4.5 ini akan membentuk pohon keputusan untuk mempermudah dalam proses pembacaan data hasil dan diharapkan akan membantu pihak yayasan dalam seleksi penerima beasiswa yayasan untuk siswa baru. dan hasil yang didapatkan sebesar 57% siswa yang dicalonkan memiliki kesempatan menerima beasiswa yang disediakan oleh SMK TI Arilangga.

8. Menurut Rosyidi(2017) penelitiannya yang berjudul **“KLASIFIKASI PENERIMAAN BANTUAN LANGSUNG SEMENTARA MASYARAKAT (BLSM) DENGAN METODE C 4.5 (Studi Kasus : Kelurahan Sembayat Kecamatan Manyar)”**. Kelurahan Sembayat Kecamatan Manyar yang menjadi salah satu daerah yang menjadi target penyaluran BLSM. Berdasarkan pengamatan di lapangan, dalam penentuan masyarakat yang menerima BLSM di Kelurahan Sembayat Kecamatan Manyar ditentukan oleh setiap ketua RT setempat, dimana yang sudah terjadi biasanya tidak objektif karena tidak mengacu pada kriteria penerima BLSM itu sendiri, sehingga masih terdapat beberapa keluarga yang tidak berhak menerima dana BLSM. Pengolahan data hasil klasifikasi masyarakat yang menerima BLSM dapat dilakukan dengan cara pengembangan sistem yang lebih efektif dalam klasifikasi, dan nantinya dapat memudahkan pegawai Kelurahan. Penelitian ini menerapkan teknik data mining klasifikasi dengan menggunakan metode decision tree C4.5 untuk menentukan kelas penerima BLSM yaitu kelas tidak terima atau kelas terima. Atribut yang digunakan pada penelitian ini terdiri dari 5 variabel, yaitu penghasilan, kondisi rumah, berat makan/hari, jumlah keluarga, dan tanggungan sekolah. Pengujian sistem dilakukan sebanyak 2 kali pengujian. Data yang digunakan diambil dari

data penerima BLSM yang terdapat pada Kelurahan Sembayat Kecamatan Manyar, sebanyak 100 data.

9. Menurut Arifianto, Rahayu(2018) penelitiannya yang berjudul **“IMPLEMENTAS I DECISSION TREE ALGORITMA C4.5 PADA KLASIFIKAS I PENDUDUK MISKIN DI KABUPATEN JEMBER”**. Di Kabupaten Jember tingkat kemiskinan masih tergolong tinggi, keterangan dari Dinas Sosial, perkembangan penduduk miskin di Kabupaten Jember dalam kurun waktu 2 tahun terakhir yakni 2016 berjumlah 265,10 ribu jiwa. Dalam pengklasifikasiannya, penduduk kurang mampu didasarkan pada beberapa parameter yaitu pendidikan, pendapatan, luas bangunan rumah, status kepemilikan rumah, jumlah makan dalam sehari, fasilitas MCK dan sumber penerangan. Untuk mengklasifikasikan data penduduk miskin dengan menggunakan 7 parameter, digunakan metode algoritma C4.5 yang mempunyai kehandalan yang baik dalam membuat klasifikasi data. Sebelum diproses menggunakan algoritma C4.5 dilakukan preprocessing data dengan mengisi missing value. Setelah dilakukan uji coba klasifikasi data penduduk miskin didapatkan tingkat akurasi tinggi sebesar 99,33%. Model ini berhasil diterapkan pada sistem untuk mengklasifikasikan data penduduk miskin yang layak untuk mendapatkan bantuan maupun tidak.
10. Menurut Oktaviani, Sarkawi(2017) penelitiannya yang berjudul **“KLASIFIKASI C4.5, NEURAL NETWORK, PSO DALAM PENENTUAN KELAYAKAN BANTUAN DANA BERGULIR”**. Koperasi kredit atau credit union adalah sebuah lembaga keuangan yang bergerak di bidang simpan pinjam yang dimiliki dan dikelola oleh anggotanya, bertujuan untuk menyejahterakan anggotanya sendiri. Kredit macet dapat dihindari dengan cara melakukan analisa kredit yang akurat terhadap calon debitur. Keakuratan penilaian kredit sangat penting untuk profitabilitas lembaga keuangan. Peningkatan akurasi penilaian kredit dapat dilakukan dengan cara melakukan seleksi terhadap atribut, karena seleksi atribut mengurangi dimensi dari data sehingga operasi algoritma data mining dapat berjalan lebih efektif dan lebih cepat. Banyak penelitian yang telah dilakukan untuk penentuan penilaian kredit. Salah satu metode yang paling banyak digunakan adalah metode algoritma C4.5 dan Neural Network. Dalam penelitian ini akan digunakan metode C4.5 dan Neural Network, akan dilakukan seleksi atribut dengan menggunakan particle swarm optimization untuk penentuan penilaian kredit. Machines berbasis particle swarm optimization masuk ke dalam kategori klasifikasi sangat baik. Penelitian ini dilakukan untuk memudahkan penentuan kredit, sehingga kredit macet dapat dihindari.

Tabel 2. 4 Tinjauan Studi

No	Peneliti	Judul	Sumber	Kontribusi
1	(Senna, 2018)	ALGORITMA KLASIFIKASI DATA MINING UNTUK MEMPREDIKSI SISWA DALAM MEMPEROLEH BANTUAN DANA PENDIDIKAN	Faktor Exacta 11(3):266-274 ,2018 p- ISSN : 1979-276X e- ISSN: 2502-339X DOI :10.30998/faktorexacta.v11i3.2777	Kontribusi: Acuan perhitungan algoritma c 4.5 dan teknik terbaik untuk Memprediksi Siswa dalam memperoleh Bantuan Dana Pendidikan.
2	(Maurina, 2020)	PENERAPAN DATA MINING UNTUK REKOMENDASI BEASISWA PADA SMA MUHAMMADIYAH GUBUG MENGGUNAKAN ALGORITMA C4.5	<a href="http://eprints.dinus.ac.id/16497/1/jurnal_15440.pdf">http://eprints.dinus.ac.id/16497/1/jurnal_15440.pdf</a>	Kontribusi: Pada penelitian ini Data yang digunakan yaitu data jurusan, kelas, jumlah nilai, penghasilan orangtua, dan jumlah saudara kandung. Untuk membuktikan bahwa algoritma C4.5 cukup akurat dalam menentukan rekomendasi beasiswa pada SMA Muhammadiyah Gubug.
3	(Isyriyah, 2016)	PERFORMANCE METODE DECISION TREE C 4.5 PADA	<a href="http://repository.stiki.ac.id/380/1/Performance%20Metode%20Decision%20Tree%20C45%20Penerimaan%20Beasiswa.pdf">http://repository.stiki.ac.id/380/1/Performance%20Metode%20Decision%20Tree%20C45%20Penerimaan%20Beasiswa.pdf</a>	Kontribusi : akurasi pengujian dengan menggunakan



No	Peneliti	Judul	Sumber	Kontribusi
		ANALISA PENGAMBILAN KEPUTUSAN PENERIMA BEASISWA		algoritma C4.5 masih baik dan dapat dijadikan salah satu pedoman untuk seleksi penerima beasiswa sebelum ada keputusan penerima beasiswa.
4	(Badriah, Nugroho, Sanjaya, Rismawati, & Rozikin, 2021)	KLASIFIKASI ALGORITMA C4. 5 DALAM MENENTUKAN PENERIMA BANTUAN COVID- 19(Studi Kasus : Desa di Karawang)	JIP (Jurnal Informatika Polinema) ISSN: 2614-6371 E-ISSN: 2407- 070X	Kontribusi: Pengambilan variabel jumlah anggota keluarga untuk dijadikan variabel yang akan digunakan.
5	(Amalianisa, 2020)	PEMILIHAN CALON PENERIMA BANTUAN SISWA MISKIN MENGUNAKAN ALGORITMA C4.5 DI SMAN 1 AEK KUASAN	<a href="http://repository.stmikroyal.ac.id/244/">http://repository.stmikroyal.ac.id /244/</a>	Kontribusi: Kontribusi terhadap penelitian ini adalah penggunaan metode algoritma c 4.5
6	(Merdekawati & Rahayu, 2021)	SISTEM PENDUKUNG KEPUTUSAN PEMBERIAN BANTUAN EKONOMI MENGUNAKAN ALGORITMA C	Serambi Engineering, Volume VI, No. 1, Januari 2021 hal 1452 - 1464 -ISSN : 2528-3561e-ISSN : 2541-1934	Kontribusi: Pengambilan variabel penghasilan untuk dijadikan variabel yang akan digunakan.

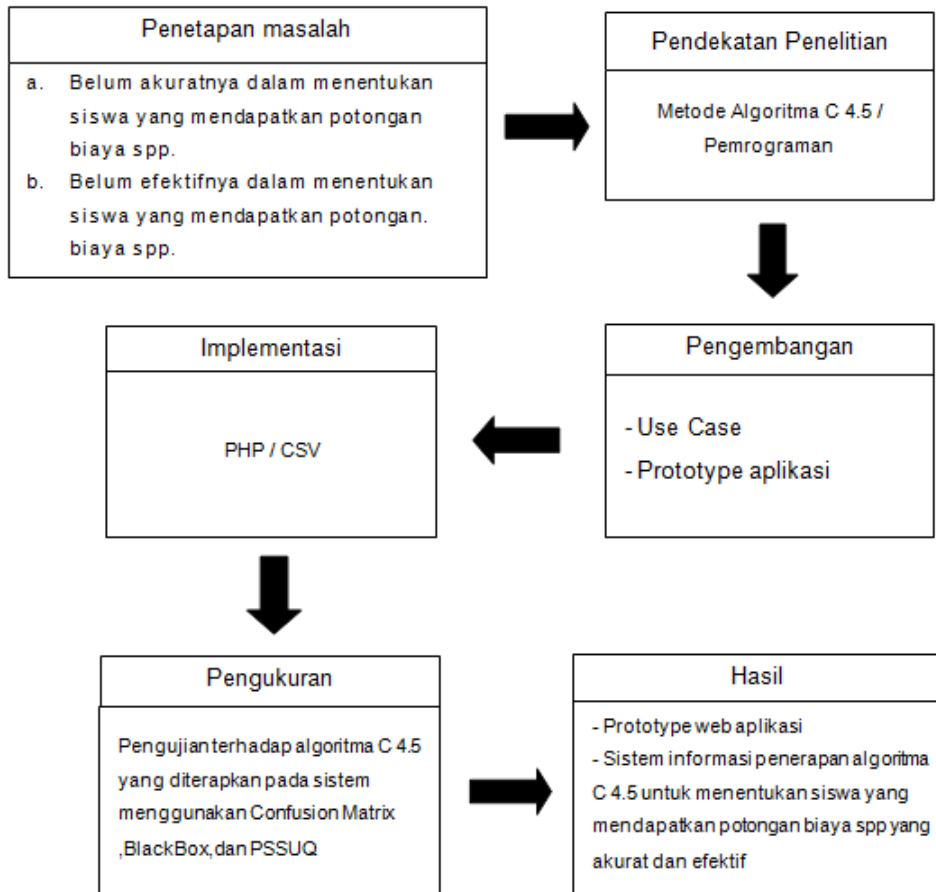
No	Peneliti	Judul	Sumber	Kontribusi
		4.5		
7	(Azhari & Nursobah, 2021)	Rekomendasi Penerimaan Beasiswa Yayasan Untuk Siswa Baru SMK TI Airlangga dengan Algoritma C4.5	Jurnal media informatika bududarma Volume 5, Nomor 2, April 2021, Page 609-614	Kontribusi: hasil dari keseluruhan penelitian menerangkan bahwasannya algoritma C4.5 mampu mengatasi permasalahan klasifikasi terhadap pembentukan pola terstruktur dalam proses penemuan informasi dan pengambilan keputusan.
8	(Rosyidi, 2017)	KLASIFIKASI PENERIMAAN BANTUAN LANGSUNG SEMENTARA MASYARAKAT (BLSM) DENGAN METODE C 4.5 (Studi Kasus : Kelurahan Sembayat Kecamatan	<a href="http://eprints.umg.ac.id/2041/1/9.%20Intisari.pdf">http://eprints.umg.ac.id/2041/1/9.%20Intisari.pdf</a>	Kontribusi: Kontribusi terhadap penelitian ini adalah penggunaan metode algoritma c 4.5

No	Peneliti	Judul	Sumber	Kontribusi
		Manyar)		
9	(Afrianto & Rahayu, 2018)	IMPLEMENTASI DECISION TREE ALGORITMA C4.5 PADA KLASIFIKASI PENDUDUK MISKIN DI KABUPATEN JEMBER	<a href="http://repository.unmuhjember.ac.id/3413/1/a.%20PENDAHULUAN.pdf">http://repository.unmuhjember.ac.id/3413/1/a.%20PENDAHULUAN.pdf</a>	Kontribusi: Model ini berhasil diterapkan pada sistem untuk mengklasifikasikan data penduduk miskin yang layak untuk mendapatkan bantuan maupun tidak.
10	(Oktaviani & Sarkawi, 2017)	KLASIFIKASI C4.5, NEURAL NETWORK, PSO DALAM PENENTUAN KELAYAKAN BANTUAN DANA BERGULIR	Konferensi Nasional Ilmu Sosial & Teknologi (KNiST) Maret 2017, pp. 119~124	Kontribusi : Dari hasil penelitian yang dilakukan maka didapat beberapa kesimpulan dengan Algoritma C4.5 dan Neural Network berbasis PSO Dapat mengklasifikasikan data dari syarat syarat sesuai dengan yang diinginkan oleh pemberi bantuan.

Metode dan hasil dari 10 jurnal tersebut menggunakan metode algoritma C 4.5 yang dapat digunakan untuk klasifikasi. Sedangkan untuk persamaan dengan penelitian ini yaitu menggunakan metode algoritma C 4.5. Dan untuk perbedaannya dengan penelitian ini yaitu metode algoritma C 4.5 dapat digunakan untuk menentukan siswa yang mendapatkan potongan biaya SPP.

### C. Kerangka Pemikiran

Berdasarkan dukungan landasan teoritis yang diperoleh dari eksplorasi teori yang dijadikan rujukan penelitian, maka dapat disusun kerangka pemikiran sebagai berikut :



Gambar 2. 6 Kerangka pemikiran

Keterangan :

Penelitian ini diawali dengan munculnya permasalahan terkait melalui pengumpulan data dengan cara observasi & wawancara yang dilakukan pada objek permasalahan.

1. Penetapan Permasalahan dalam penelitian ini yaitu belum akuratnya dalam menentukan siswa yang mendapatkan potongan biaya spp dan belum efektifnya dalam menentukan siswa yang mendapatkan potongan biaya spp .

2. Pendekatan penelitian ini yaitu, untuk menentukan siswa yang mendapatkan potongan biaya spp. Metode penelitian pengembangan ini menggunakan metode algoritma C.4.5 sebagai dasar analisis data penelitian dan
3. pengembangan yang digunakan dalam pembangunan system nantinya menggunakan use case dan aplikasi prototype.
4. implementasi sebagai dasar evaluasi dan validasi system prototype yang akan dibuat nantinya menggunakan php / csv.
5. Pengukuran pengujian terhadap algoritma C4.5 yang diterapkan pada sistem menggunakan Confusion Matrix, Blackbox dan PSSUQ
6. Hasil yang diharapkan dari penelitian ini yaitu menghasilkan system informasi prototype berbasis website khusus sebagai informasi terkait siswa yang mendapatkan potongan biaya spp yang lebih akurat dan efektif.

#### **D. Hipotesis Penelitian**

Hipotesis adalah jawaban sementara terhadap masalah yang masih bersifat praduga karena masih harus dibuktikan kebenarannya. algoritma C4.5, yaitu algoritma decision tree atau membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi yang terkenal. Pohon keputusan berguna sekali dalam mengeksplorasi data-data, menemukan hubungan tersembunyi antara sejumlah calon variable input dengan sebuah variabel target. Penelitian yang dilakukan sebelumnya oleh Dina Maurina(2016) "PENERAPAN DATA MINING UNTUK REKOMENDASI BEASISWA PADA SMA MUHAMMADIYAH GUBUG MENGGUNAKAN ALGORITMA C4.5". Maka hipotesis dalam penelitian ini penerapan metode algoritma c 4.5 diduga dapat menentukan siswa yang mendapatkan potongan biaya spp. Maka diharapkan pemberian potongan biaya spp tersebut tepat sasaran dan sesuai target, yang mana bantuan tersebut benar-benar diterima oleh siswa yang membutuhkan.