

BAB II

KERANGKA TEORITIS

A. Landasan Teori

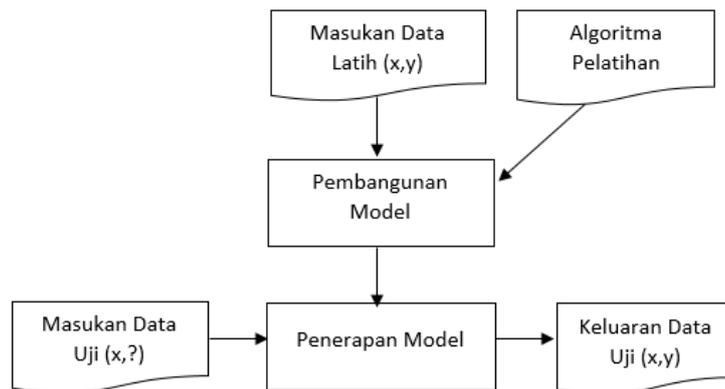
1. Data Mining

Menurut (Vulandari, 2017, hal. 1) menyatakan bahwa data mining artinya proses menganalisis untuk menemukan pengetahuan yang disimpan pada *database*; informasi yang dihasilkan didapatkan dengan mengekstraksi dan mengenali pola yang penting dan menarik dari data yang terdapat pada basis data; data mining terutama digunakan dalam menggali pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut dengan *Knowledge Discovery In Database* yang disingkat KDD; proses KDD memiliki beberapa tahapan di antaranya adalah :

- (1) Data Cleansing, yaitu proses pembersihan data untuk memilih data yang dianggap bisa dipakai atau melepaskan data yang tidak selaras (data yang tidak konsisten dan *noise*);
- (2) Data Integration, yaitu proses menggabungkan data yang dianggap berulang akan digabungkan menjadi satu atau kombinasi data dari beberapa sumber;
- (3) Data Selection, Pemilihan data yaitu proses seleksi atau pemilihan data yang dianggap relevan terhadap analisis.
- (4) Data Transformation, yaitu proses transformasi data terpilih ke dalam bentuk mining procedure atau merubah data menjadi bentuk yang sesuai agar dapat di mining;
- (5) Data Mining, yaitu proses dimana dilakukan beragam teknik untuk mengekstrak pola-pola potensial menghasilkan data yang berguna atau proses ekstraksi pola dari data yang ada;
- (6) Pattern Evolution, yaitu proses dimana pola-pola yang telah diidentifikasi berdasarkan measure yang diberikan atau menafsirkan pola menjadi pengetahuan yang dapat digunakan untuk mengambil keputusan;
- (7) Knowledge Presentation, yaitu data yang sudah diproses divisualisasikan ke dalam bentuk gambar, tulisan, grafik, dan lain-lain agar lebih mudah dipahami oleh pengguna dan diharapkan bisa diambil tindakan berdasarkan analisis.

2. Klasifikasi

Menurut (Prasetyo, 2013:45) menyatakan bahwa klasifikasi adalah aktivitas memperkirakan data untuk menggolongkan ke dalam golongan yang telah ditetapkan asal sejumlah golongan yang sudah ada; rancangan kerja ditunjukkan pada Gambar 2.1 dimana memiliki beberapa data latih (x,y) berfungsi untuk data pembangun pola berdasarkan data latih kemudian berfungsi untuk memperkirakan kelas dari data uji (x,?) sehingga diketahui kelas y yang sebenarnya;



**Gambar 2. 1 Proses Pekerjaan Klasifikasi
(Sumber : Prasetyo, 2013)**

Pola yang sudah dibuat ketika pelatihan kemudian dapat digunakan untuk memprediksi label kelas data baru yang belum diketahui; dalam pembangunan pola selama proses pelatihan memerlukan suatu algoritma untuk membangunnya yaitu Algoritma pelatihan; ada berbagai macam algoritma pelatihan yang sudah dikembangkan oleh para peneliti yaitu C4.5, Nearest Neighbor, Bayesian Classification, Neural Network, dan sebagainya.

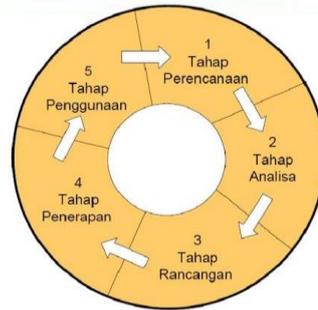
3. Pengembangan Sistem SDLC

Pendekatan sistem merupakan sebuah metodologi; metodologi adalah aturan yang direkomendasikan untuk melakukan sesuatu; pendekatan sistem adalah metodologi dasar dalam menyelesaikan berbagai macam masalah; siklus hidup pengembangan sistem (*System Development Life Cycle – SDLC*) adalah penerapan sebuah metode untuk mengembangkan suatu sistem informasi (Everett & McLeod, 2006, p.200) ada berbagai macam urutan dalam pengembangan apabila proyek tersebut kemungkinan akan berhasil; Urutan yang dapat dilakukan sebagai berikut :

- (1) Merencanakan;
- (2) Melakukan analisis;
- (3) Pembuatan desain;
- (4) Melakukan implementasi;
- (5) Pemakaian;

Sebelum melaksanakan proyek, maka perlu merencanakan sumber daya apa saja yang diperlukan; melakukan analisis terhadap sistem yang sudah serta menentukan permasalahan dan kebutuhan pada sistem baru yang akan dikembangkan; kemudian merancang dan menerapkannya terhadap sistem baru; kemudian sistem baru tersebut digunakan dan diharapkan dapat bertahan dalam waktu yang cukup panjang; karena urutan di atas mengikuti struktur yang sistematis dan dikerjakan dengan cara *top-down*, umumnya proses air terjun (*waterfall approach*) sering

disebut sebagai SDLC tradisional; proses ini mempunyai aliran satu arah dalam menyelesaikan proyek sebagai berikut;



Gambar 2. 2 Pola Melingkar dari Siklus Hidup Sistem

Sumber: (Everett & McLeod, 2006, p.200)

Gambar 2.2 menggambarkan siklus hidup dengan sifat melingkar; saat sebuah sistem yang digunakan telah memiliki banyak kekurangan, maka harus dikembangkan sistem baru yang menggunakan siklus hidup baru dan dimulai dari tahap perencanaan; permasalahan dapat ditentukan pada bagian merencanakan dan menganalisa; cara menyelesaikan masalah tersebut kemudian ditetapkan dan dinilai pada bagian desain; lalu, cara terbaik diterapkan; selama bagian pemakaian, hal tersebut dilakukan untuk mengetahui seberapa baik sistem baru untuk memecahkan permasalahan yang telah ditetapkan.

4. Metode Prototype

Menurut (Ogedebe & Jacob, 2012, hal. 219) menjelaskan *prototyping* adalah cara mengembangkan sistem, yaitu model alur proses pada sebuah sistem dan berfungsi sebagai tampilan awal pada sistem yang akan dikembangkan yaitu :

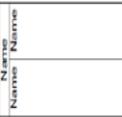
- (a) Illustrative, mengeluarkan tampilan antarmuka sistem;
 - (b) Simulated, penggambaran beberapa alur kerja sistem namun belum data aktual;
 - (c) Functional, penggambaran beberapa alur kerja sistem sesuai data real;
 - (d) Evolutionary, memperoleh pola yang digunakan sebagai elemen dari alur sistem;
- manfaat dari penggunaan *prototyping* :

- (a) Menjadikan sistem sebenarnya dalam tiruan sistem yang akan dikembangkan, serta demi kesempurnaan sistem maka perlukan juga saran dari pengguna;
- (b) Setiap perubahan sampai hasil akhir pada sistem yang akan dikembangkan yang disesuaikan dengan *prototype*, pengguna akan lebih siap menerima perubahan-perubahan yang terjadi;
- (c) Pengembangan *prototype* disesuaikan dengan pengguna, kemungkinan yang terjadi yaitu *prototype* akan mengalami penambahan ataupun pengurangan dalam proses pengembangannya;
- (d) Demi menghasilkan sistem informasi yang tepat dan lebih baik bagi pengguna maka meminimalisir penggunaan sumber daya dan waktu.

5. Business Process Model and Notation (BPMN)

Business Process Model and Notation (BPMN) merupakan sebuah representasi grafis dalam model proses bisnis untuk menentukan proses bisnis; tujuan BPMN untuk menyediakan pengguna teknis dan bisnis dengan markup intuitif kepada pelanggan bisnis manajemen proses bisnis, mengekspresikan kualitas dari proses yang kompleks; Berikut simbol BPMN (Pane, Lase, & Mali, 2020) :

Simbol BPMN

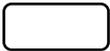
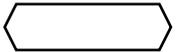
SIMBOL	NAMA	KETERANGAN
	Start	Merupakan awal dimulainya suatu aktivitas;
	Intermediate	Terjadi setelah proses sudah dimulai dan sebelum suatu proses berakhir;
	End	Merupakan akhir dari segala aktivitas;
	Pool	Digunakan sebagai tempat grafis pada partisi satu set proses dari pool lain;
	Lane	Digunakan untuk mengidentifikasi aktor yang terlibat;
	Abstract	Aktivitas yang dilakukan;
	User Task	Aktivitas yang dilakukan oleh user dengan menggunakan perangkat lunak;

SIMBOL	NAMA	KETERANGAN
	Script Task	Ketika task atau aktivitas dimulai maka mesin akan menjalankan script, begitu pula sebaliknya;
	Manual Task	Aktivitas yang dijalankan tanpa bantuan mesin atau aplikasi apapun;
	Business Rule Task	Aktivitas akan memungkinkan mengirimkan data ke dan menerima data dari business rule engine;
	Service Task	Aktivitas yang dijalankan akan otomatis oleh aplikasi;
	Exclusive Gateway	Kondisi dimana hanya ada satu pilihan di dalam nya;
	Parallel Gateway	Kondisi dimana adanya beberapa pilihan yang dilalui;
	Inclusive Gateway	Kondisi dimana adanya satu kondisi atau lebih bisa dilalui;
	Sequence Flow	Penghubung untuk task berikutnya yang terdapat pada satu line;
	Message Flow	Digunakan untuk menyampaikan pesan dari dua pool;

SIMBOL	NAMA	KETERANGAN
----- ;	Association ; Flow ;	Digunakan untuk menghubungkan elemen dengan artifact.

6. Flowchart

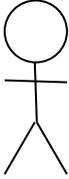
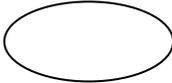
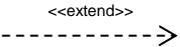
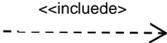
Menurut (Setiawan 2009:25) menyatakan bahwa flowchart atau bagan alur adalah diagram yang menampilkan langkah-langkah dan keputusan untuk melakukan sebuah proses dari suatu program; setiap langkah digambarkan dalam bentuk diagram dan dihubungkan dengan garis atau arah panah; dengan simbol – simbol :

Simbol Flowchart	
SIMBOL	Fungsi
	; Permulaan sub program;
	; Perbandingan, pernyataan; penyeleksian data yang memberikan pilihan untuk langkah selanjutnya;
	; Penghubung bagian bagian flowchart yang berada pada satu halaman;
	; Penghubung bagian bagian flowchart yang berada pada halaman berbeda;
	; Permulaan/akhir program;
	; Arah aliran program;
	; Proses inisialisasi/pemberian harga awal;
	; Proses penghitung/ proses pengolahan data;
	; Proses input atau output data.

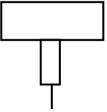
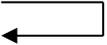
7. Unified Modeling Language (UML)

Menurut (Dharwiyanti, 2003, hal. 4) menyatakan bahwa *Unified Modeling Language* (UML) adalah sebuah gambar yang bertujuan untuk mengungkapkan gagasan dalam bentuk simbol, grafik, tulisan, dan lain lain pada sistem yang akan dibangun; dengan simbol – simbol sebagai berikut :

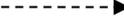
Simbol Use Case Diagram

SIMBOL	NAMA	KETERANGAN
	; <i>ACTOR</i> ;	Simbol yang digunakan untuk menjelaskan siapa saja pengguna yang terdapat pada <i>system</i> ;
	; <i>USE CASE</i> ;	Aktivitas yang dilakukan oleh pengguna sistem, biasanya menggunakan awalan kata kerja;
	<i>ASOSIASI</i> / ; <i>ASSOCIATION</i> ;	Relasi yang digunakan untuk interaksi antar use case dan actor;
	; <i>EKSTENSI</i> / ; <i>EXTEND</i> ;	Relasi yang digunakan untuk interaksi use case tambahan dengan use case. Meskipun use case tambahan tersebut mampu berdiri sendiri;
	; <i>GENERALISASI</i> / ; <i>GENERALIZATI</i> / ; <i>ON</i> ;	Hubungan generalisasi dari dua use case, di mana satu fungsi lebih umum daripada yang lain;
	; <i>MENGGUNAKAN</i> / ; <i>INCLUDE</i> ;	Relasi usecase pelengkap dengan usecase lain, yaitu usecase pelengkap tersebut memerlukan use case lain untuk menjalankan pekerjaannya.

Simbol Sequence Diagram

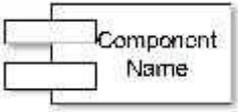
SIMBOL	NAMA	KETERANGAN
	<i>Actor</i>	Simbol siapa saja yang dapat menjalankan fungsi pada sistem;
	<i>Lifeline</i>	Simbol yang menghubungkan antar objek pada sequence;
	<i>General</i>	Merepresentasikan entitas tunggal dalam sequence;
	<i>Boundary</i>	Simbol yang digunakan sebagai halaman tampilan pada sistem;
	<i>Control</i>	Simbol yang digunakan sebagai proses pada fungsi yang dijalankan pada sistem;
	<i>Entitas</i>	Simbol yang digunakan untuk menyimpan data atau informasi setelah fungsi pada sistem dijalankan;
	<i>Activation</i>	Simbol yang digunakan untuk menunjukkan sebuah objek mengirim atau menerima objek;
	<i>Message Entry</i>	Simbol yang digunakan sebagai hubungan antar objek untuk aktivitas yang dilakukan setelahnya;
	<i>Message to Self</i>	Simbol yang digunakan untuk hubungan objek itu sendiri;
	<i>Message Return</i>	Simbol yang digunakan untuk mengirim pesan dari arah kanan ke kiri.

Simbol Class Diagram

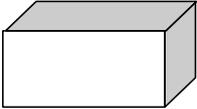
SIMBOL	NAMA	KETERANGAN
 ;	<i>Generalization</i> ;	Simbol yang menghubungkan antara objek satu dengan yang lainnya;
 ;	<i>Nary Association</i> ;	Simbol yang berisi keterangan untuk menghubungkan lebih dari 2 objek;
 ;	<i>Class</i> ;	Simbol terhadap interface, operasi sistem, dan tabel pada database;
 ;	<i>Realization</i> ;	Simbol relasi fungsi yang sebenarnya dikerjakan oleh objek tersebut;
 ;	<i>Dependency</i> ;	Simbol relasi yang bersifat mandiri pada sebuah elemen;
 ;	<i>Association</i> ;	Simbol relasi antar objek.

Simbol Component Diagram

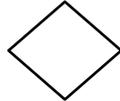
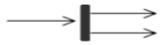
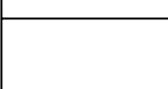
SIMBOL	NAMA	KETERANGAN
 ;	<i>Package</i> ;	Kumpulan dari beberapa Komponen;

SIMBOL	NAMA	KETERANGAN
	- , <i>Komponen</i>	- , Komponen apa saja yang ada pada system;
	- , Kebergantungan/ <i>dependency</i>	- , Penghubung yang menuju pada node yang digunakan;
	- , <i>Link</i>	- , Penghubung antara node satu dengan node lainnya.

Simbol Deployment Diagram

SIMBOL	NAMA	KETERANGAN
	- , <i>Package</i>	- , Kumpulan dari beberapa Node;
	- , <i>Node</i>	- , Merupakan simbol hardware dan software;
	- , Kebergantungan/ <i>Dependency</i>	- , Penghubung yang menuju pada node.

Simbol Activity Diagram

GAMBAR	NAMA	KETERANGAN
	<i>INITIAL</i>	Status awal dari aktivitas suatu sistem;
	<i>ACTIVITY</i>	Aktivitas yang dilakukan sistem, yang diawali dengan kata kerja;
	<i>PERCABANGAN/ DECISION</i>	Aktivitas percabangan yang melakukan aktivitas dengan menggabungkan menjadi satu aktivitas;
	<i>PENGGABUNGAN/ JOIN</i>	Aktivitas menggabungkan satu aktivitas ke aktivitas yang lain;
	<i>STATUS AKHIR/FINAL</i>	Status akhir yang dilakukan sistem, sebagai aktivitas yang memiliki sebuah status baru;
	<i>SWIMLINE</i>	Memisahkan sebuah proses bisnis terhadap aktivitas dari suatu sistem.

8. Database

Menurut (Mustakini, 2009, hal. 49) database adalah kumpulan dari data-data yang saling berhubungan satu dengan yang lainnya, tersimpan di perangkat keras komputer dan digunakan perangkat lunak untuk memanipulasi; Basis data merupakan salah satu komponen yang penting dalam sistem informasi, karena merupakan basis dalam menyediakan informasi bagi para pemakai; Perancangan database yang digunakan adalah untuk memudahkan dalam mengetahui file-file database yang digunakan dalam perancangan sistem, sekaligus untuk mengetahui hubungan antara file dari database tersebut.

9. Webserver (Apache)

Menurut (Kurniawan, 2008, hal. 2), menjelaskan bahwa *webserver* 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 adalah apache dan microsoft internet information service (IIS); dengan ilustrasi cara kerja seperti pada Gambar 2.3



Gambar 2. 3 Cara Kerja Web Server

Sumber : (Kurniawan, 2008, hal. 2)

- (1) Web server merupakan mesin dimana tempat aplikasi atau software beroperasi dalam mendistribusikan web *page* ke *user*, tentu saja sesuai dengan permintaan *user* ;
- (2) Hubungan antara web server dan browser internet merupakan gabungan atau jaringan komputer yang ada di seluruh dunia. Setelah terhubung secara fisik, protocol TCP/IP (*networking protocol*) yang memungkinkan semua komputer dapat berkomunikasi satu dengan yang lainnya; pada saat *browser* meminta data web *page* ke server maka instruksi permintaan data oleh browser tersebut di kemas di dalam TCP yg merupakan protocol transport dan dikirim ke alamat yg dalam hal ini merupakan protocol berikutnya yaitu *Hyper Text Transfer Protocol (HTTP)*.

10. Pemrograman

Penelitian ini menggunakan PHP dalam pemrogramannya. Menurut (Setiawan, 2017, hal. 54-59) PHP sendiri sebenarnya merupakan singkatan dari "*Hypertext Preprocessor*", yang merupakan bahasa scripting tingkat tinggi yang dipasang pada dokumen HTML; sebagian besar sintaks dalam PHP mirip dengan bahasa C, Java, dan Perl, namun pada PHP ada beberapa fungsi yang lebih spesifik; sedangkan tujuan utama dari penggunaan bahasa ini adalah untuk memungkinkan perancang web yang dinamis dan dapat bekerja secara otomatis; untuk membuat halaman web, sebenarnya PHP bukanlah bahasa pemrograman yang wajib digunakan; kita bisa saja membuat website hanya menggunakan HTML saja; web yang dihasilkan dengan HTML (dan CSS) ini dikenal dengan website statis, di mana konten dan halaman web bersifat tetap; sebagai perbandingan, website dinamis yang bisa dibuat menggunakan PHP adalah situs web yang bisa menyesuaikan tampilan konten tergantung situasi; website dinamis juga bisa menyimpan data ke dalam database, membuat halaman yang berubah-ubah sesuai input dari user, memproses form, dan lain sebagainya.

Dalam penelitian ini bahasa pemrograman PHP akan digunakan dalam mengembangkan sebuah sistem berbasis website, yang nanti sistem tersebut akan berbentuk menjadi website yang dinamis, karena web dinamis sendiri cenderung lebih mudah digunakan karena mendukung perubahan informasi langsung oleh

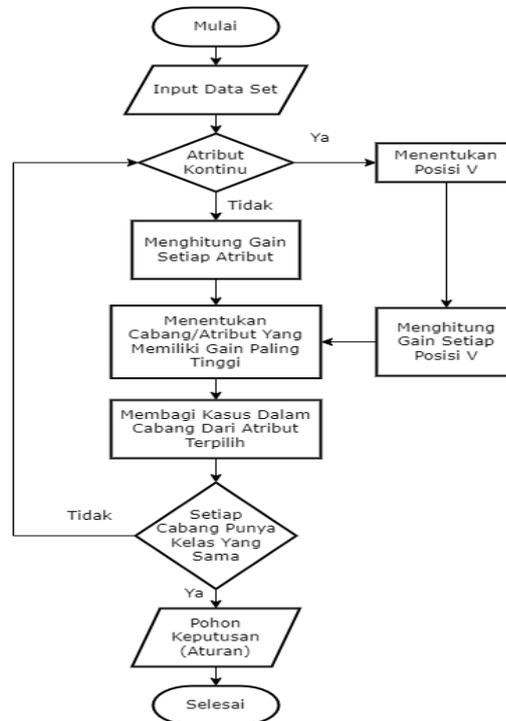
pengguna. Juga PHP sendiri akan berperan penting dalam pengambilan data yang berada dalam database, sehingga bentuk informasi dalam website tidak berupa statis. PHP bisa diartikan sebagai bahasa untuk *backend*, karena sistem bekerja di balik layar untuk mengolah database dan juga server.

B. Metode Algoritma C4.5

Menurut (Nofriansyah 2015:57) Algoritma C4.5 suatu metode algoritma membentuk berdasarkan kriteria-kriteria pembentukan keputusan; pohon keputusan adalah sebuah metode klasifikasi dan prediksi yang terkenal dan kuat; metode pohon keputusan dengan mempresentasikan aturan dan mengubah sebuah fakta yang besar menjadi pohon keputusan; pohon keputusan berguna juga untuk mengeksplorasi data, menginput sebuah target variable bahwa menemukan hubungan tersembunyi antara sejumlah variable yang diinput; pohon keputusan dapat digunakan untuk membagi sebuah kumpulan data besar menjadi beberapa himpunan-himpunan record kecil dengan cara menerapkan aturan keputusan; berikut adalah contoh kasusnya; di salah satu universitas di butuhkan sebuah alat analisis berupa sebuah aplikasi penunjang keputusan berbasis komputer untuk menentukan kelayakan mahasiswa yang akan mengikuti semester pendek; proses dimulai dari mengupulkan data mahasiswa yang akan di uji, kemudian akan di lakukan langkah langkah data preprocessing yaitu :

- (1) *Data selection* dalam proses ini data akan dipilih mana yang akan berguna untuk dipakai sebagai atribut
- (2) *Data cleaning* adalah proses mengisi atribut yang hilang atau kosong, dan merubah data yang tidak konsisten;
- (3) *Data transformation* dalam proses ini data ditransformasikan kedalam bentuk yang sesuai untuk keperluan data mining;
- (4) *Data reduction* adalah proses memperkecil ukuran database dengan menghilangkan atribut yang tidak terpakai dan menyisakan atribut yang diperlukan untuk proses data mining.

Berikut algoritma C4.5 menggunakan flowchart yang disajikan beserta penjelasannya :



Gambar 2. 4 Flowchart algoritma C4.5

Sumber : (Larose, 2005)

- (a) Memasukan Data Set yang telah disediakan; Data Set adalah kumpulan objek dan atributnya;
- (b) pada proses cek keputusan dilihat apakah data set tersebut termasuk atribut kontinu atau bukan, atribut kontinu adalah atribut yang mempunyai jangkauan real;
- (c) jika Iya maka akan menentukan posisi V, setelah itu menghitung gain setiap V; jika tidak maka menghitung gain setiap atribut;
- (d) setelah itu menentukan cabang/atribut yang memiliki gain paling tinggi;
- (e) setelah itu membagi kasus dalam cabang dari atribut yang terpilih;
- (f) jika setiap cabang mempunyai kelas yang sama, maka menghasilkan pohon keputusan; jika tidak maka cabang atau atribut tersebut kembali melakukan cek atribut kontinu.

C. Penentuan Calon Peserta Lomba Kompetisi Sains Nasional (KSN)

Menurut (KBBI, 2016) penentuan dapat dimaknai sebagai proses, cara, perbuatan menentukan; penetapan; pembatasan (arti dan sebagainya); calon dapat diartikan sebagai orang yang akan menjadi; peserta dapat diartikan orang yang ikut serta atau yang mengambil bagian (misal dalam kongres, seminar, lokakarya, dan pertandingan). Lomba Kompetisi Sains Nasional (KSN) adalah ajang Kompetisi Sains bagi pelajar SD, SMP, dan SMA di seluruh Indonesia yang diselenggarakan oleh Pusat Prestasi Nasional. Sebelum berganti nama di tahun 2020, KSN dikenal dengan

nama Olimpiade Sains Nasional (OSN). Perubahan ini terjadi setelah Menteri Pendidikan, Kebudayaan, Riset dan Teknologi Republik Indonesia, Nadiem Makarim, mendirikan Pusat Prestasi Nasional yang menaungi seluruh pelaksanaan lomba, kompetisi, dan festival di Indonesia, guna meningkatkan mutu pendidikan dan memberikan ruang aspirasi serta apresiasi untuk pelajar dalam mewujudkan talenta emas Indonesia (<https://ksn.puspresnas.id/>).

Jadi, penentuan calon peserta lomba Kompetisi Sains Nasional (KSN) adalah suatu proses dalam menentukan siswa yang layak untuk mengikuti lomba Kompetisi Sains Nasional (KSN) yang diselenggarakan oleh Menteri Pendidikan, Kebudayaan, Riset dan Teknologi Republik Indonesia.

D. Tinjauan Pustaka

Penelitian rujukan yang digunakan untuk penelitian ini, yaitu :

- 1. Algoritma C4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa SMP dengan menggunakan aplikasi rapid miner** (Ardiansyah & Walim, 2018). Data yang digunakan dalam penelitian ini adalah data berdasarkan kriteria yang digunakan dalam perhitungan, yaitu pada siswa SMP kelas VIII yang digunakan dalam perhitungan alternatif tertinggi penentuan siswa yang akan mengikuti lomba cerdas cermat. Metode yang diusulkan untuk proses seperti yang digunakan algoritma C4.5 dengan kriteria yang digunakan sebagai berikut: Nama Siswa, Nilai Bahasa Indonesia, Nilai Bahasa Inggris, Nilai IPA, Nilai IPS, Nilai Matematika, Nilai keaktifan, Perolehan Skor IQ dan nilai bimbingan belajar. (1) Klasifikasi proses seleksi calon peserta lomba siswa SMP dapat mengklasifikasikan siswa dalam tahapan lolos atau tidaknya dalam seleksi. (2) Dari 33 data siswa yang digunakan menunjukkan tingkat akurasi dengan algoritma C4.5 sebesar 81,81%.
- 2. Klasifikasi siswa yang akan mengikuti lomba Olimpiade Sains Nasional (OSN) menggunakan algoritma C4.5** (Fauzi, 2018). Data yang digunakan dalam penelitian ini adalah data berdasarkan kriteria yang digunakan dalam perhitungan, yaitu pada siswa SMPN 2 GRESIK kelas 2 A, B dan C semester ganjil. Metode yang digunakan adalah metode klasifikasi dengan Algoritma C4.5 dengan kriteria yang digunakan sebagai berikut : Nama siswa, Nilai ulangan harian, UTS , UAS (IPA, IPS, Matematika, Bahasa inggris). (1) Sistem ini dapat menghasilkan informasi untuk rekomendasi siswa yang akan terpilih mengikuti lomba Olimpiade Sains Nasional menggunakan metode Decision Tree C4.5. (2) Pemilihan data training dalam pengujian ini menggunakan trial and error. Dari tiga kali percobaan, pohon keputusan terbaik yang digunakan untuk mengklasifikasi status siswa adalah pohon keputusan pada percobaan ke-2. (3) Komposisi data training akan mempengaruhi tingkat akurasi yang didapatkan. Dimana setiap percobaan yang telah dilakukan akan mendapatkan akurasi yang berbeda-beda dan membentuk pohon keputusan yang terbaik.

- 3. Penerapan Algoritma C4.5 Dalam Memprediksi Kesiapan Siswa SMP IT PAPB Semarang Menghadapi Ujian Nasional** (Rahma & Setyono, 2018). Data yang digunakan dalam penelitian ini adalah data berdasarkan kriteria yang digunakan dalam perhitungan, yaitu hasil try out siswa kelas IX SMP Islam Terpadu PAPB Semarang tahun ajaran 2013/2014 – 2015/2016. Metode yang digunakan adalah metode algoritma C4.5 dengan kriteria yang digunakan sebagai berikut : Nama siswa, Kelas, NIS, Nilai Tryout (Bahasa Indonesia, Matematika, Bahasa Inggris, IPA). (1) Metode klasifikasi dapat diterapkan untuk memprediksi kesiapan siswa SMP Islam Terpadu PAPB Semarang menghadapi ujian nasional. Hal tersebut berdasarkan pengujian yang telah dijelaskan pada bab sebelumnya, pemodelan algoritma C4.5 menghasilkan akurasi sebesar 99,48%. (2) Penelitian ini dapat memudahkan pihak sekolah dalam memprediksi kesiapan siswa menghadapi ujian nasional dengan cara menyebar kuesioner kepada pihak sekolah mengenai aplikasi yang telah dibuat penulis menggunakan Java. Setelah melakukan penyebaran kuesioner tersebut memperoleh hasil presentase sebesar 83,3% yang berarti pihak sekolah puas terhadap aplikasi yang telah dibuat.
- 4. Rancangan Aplikasi Prediksi Kelulusan Siswa Menggunakan Algoritma C4.5** (Yuningsih, Setiawan, & Sunarto, 2020). Data yang digunakan dalam penelitian ini adalah data berdasarkan kriteria yang digunakan dalam perhitungan, yaitu 104 data sampel nilai semester 1-5 siswa SMK Al- Mizab tahun 2018/2019. Metode yang digunakan adalah metode algoritma C4.5 dengan kriteria yang digunakan sebagai berikut : NIS, Data pokok (Semester 1-5, nilai sikap), Data tambahan (Tahun masuk, tahun lulus). Dari hasil uji coba, indeks prestasi gain tertinggi yaitu di semester 2 dengan nilai gain 0.31667. Aplikasi ini untuk memudahkan dalam memprediksi kelulusan siswa dengan menggunakan Rule yang dihasilkan dari aplikasi. Hasil Sistem ini hanya menjadi alat bantu bagi pengambil keputusan, keputusan akhir tetap berada ditangan pengambil keputusan. Walaupun telah menghasilkan rule yang dapat menjadi pedoman dalam mengambil suatu keputusan, rule yang dihasilkan tersebut masih perlu pengujian faktual untuk mengetahui tingkat akurasinya.
- 5. ANALISIS KINERJA ALGORITMA C4.5 DAN NAÏVE BAYES DALAM MEMPREDIKSI KEBERHASILAN SEKOLAH MENGHADAPI UN** (Angraini, Fauziah, & Putra, 2020). Data yang digunakan dalam penelitian ini adalah data berdasarkan kriteria yang digunakan dalam perhitungan, yaitu data hasil ujian nasional sekolah menengah atas di daerah banda aceh. Metode yang digunakan adalah metode algoritma C4.5 dan naive bayes dengan kriteria yang digunakan sebagai berikut : rata-rata hasil Ujian Nasional (UN) siswa setiap sekolah di Banda Aceh tahun 2012. Keseluruhan data tersedia dari seluruh sekolah, seluruh provinsi di Indonesia dan dapat diakses melalui portal Open Data Banda Aceh (OBDA) di link data.bandaacehkota.go.id. Dari hasil

pengujian dengan mengukur kinerja kedua metode tersebut menggunakan confusion matrix, kurva ROC dan t-Test pada dataset diketahui bahwa Naïve Bayes memiliki nilai akurasi 95,50% dan signifikan terhadap algoritma C4.5 memiliki nilai akurasi 78,50%. Adapun model yang telah terbentuk selanjutnya dapat dikembangkan dan dapat diimplementasikan ke dalam sebuah aplikasi sehingga dapat membantu dan memudahkan bagi para pengambil keputusan untuk memprediksi prestasi siswa.

- 6. PENERAPAN TEKNIK DATA MINING UNTUK MENENTUKAN HASIL SELEKSI MASUK SMAN 1 GIBEBER UNTUK SISWA BARU MENGGUNAKAN DECISION TREE** (Sugianto, 2019). Data yang digunakan dalam penelitian ini adalah data berdasarkan kriteria yang digunakan dalam perhitungan, yaitu SMA Negeri 1 Cibeer tahun 2011 dan tahun 2012. Metode yang digunakan adalah pengujian pada RapidMiner terhadap 3 (tiga) jenis Algoritma, yaitu Algoritma C4.5, naïve Bayes. dan Neural Network dengan kriteria yang digunakan sebagai berikut : Nama Calon, Nilai UN (Bahasa Indonesia, Bahasa Inggris, Matematika, IPA, Jumlah UN, Prestasi). Tingkat akurasi penentuan prediksi siswa yang diterima cukup tinggi yaitu mencapai sebesar 99.05%.
- 7. Penerapan Algoritma C4.5 dalam Menentukan Kandidat Siswa/Siswi SMK dalam Mengikuti Lomba Kompetensi Siswa** (Putra, 2016). Data yang digunakan dalam penelitian ini adalah data berdasarkan kriteria yang digunakan dalam perhitungan, yaitu Siswa kelas XIII SMK Nurul Amaliyah Tanjung Morawa – Deli Serdang. Metode yang digunakan adalah metode algoritma C4.5 dengan kriteria yang digunakan sebagai berikut : NIS, Nama, Nilai pelajaran Bahasa Inggris, Nilai Menyusun Laporan Keuangan, Nilai Komputer Akuntansi (Spreadsheet). (1) Berdasarkan analisa yang dilakukan, untuk proses penerapan algoritma C4.5 maka terlebih dahulu untuk menentukan kriteria yang diperlukan untuk proses pemilihan kandidat siswa yang nanti akan dipilih untuk mengikuti Lomba Kompetensi Siswa. (2) Untuk membuat suatu pola keputusan yang dapat dijadikan landasan, maka harus melakukan proses penambangan data pemilihan kandidat siswa pada Sekolah Nurul Amaliyah Tanjung Morawa untuk perhitungan dengan menggunakan rumus C4.5. (3) Proses pengimplementasian algoritma C4.5 untuk mendapatkan suatu pohon keputusan dapat diuji dengan menggunakan aplikasi rapid miner berdasarkan data yang sudah ada maupun data baru.
- 8. IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL K-NEAREST NEIGHBOR, DECISION TREE SERTA NAIVE BAYES** (Hozairi, Anwari, & Alim, 2021). Data yang digunakan dalam penelitian ini adalah data berdasarkan kriteria yang digunakan dalam perhitungan yaitu mahasiswa Prodi Teknik Informatika Universitas Islam Madura dengan menerapkan Orange. Metode yang digunakan adalah metode KNN (K-Nearest Neighbor), Naive Bayes dan Decision Tree algoritma C4.5 dengan kriteria

yang digunakan sebagai berikut : NIM, Nama, Jenis Kelamin, Usia, IP Sementara semester 1-8. Diperoleh hasil bahwa kinerja Naive Bayes lebih unggul dari K-Nearest Neighbor serta Decision Tree. Terbukti bahwa dari 35 data uji yang digunakan Naive Bayes memiliki nilai akurasi 89%, presisi 88% sedangkan K-Nearest Neighbor memiliki nilai akurasi 77% , presisi 76% dan Decision Tree memiliki nilai akurasi 74% dan presisi 84%.

9. Sistem Informasi Pendukung Keputusan Penerimaan dan Monitoring Beasiswa Menggunakan Algoritma C4.5 (Studi Kasus Universitas Negeri Surabaya) (Rofik & Prehanto, 2020).

Data yang digunakan dalam penelitian ini adalah data berdasarkan kriteria yang digunakan dalam perhitungan, yaitu Mahasiswa Universitas Negeri Surabaya dengan Ipk Mahasiswa harus diatas 3, Batas masa studi mahasiswa S1 dibawah 9 Semester dan D3 dibawah 5 Semester, Memiliki prestasi Non – Akademik : Untuk D3 minimal nilai prestasi akademik adalah di atas 300, Untuk S1 minimal nilai prestasi akademik adalah diatas 500. Metode yang digunakan adalah metode algoritma C4.5 dengan kriteria yang digunakan sebagai berikut : NIM, Nama, Lama Semester, IPK, Prestasi non-akademik. Berdasarkan masing – masing karakteristik beasiswa yang ada, hasil pada penelitian ini diperoleh tingkat akurasi sebesar 81,67% dengan cara mengklasifikasikan beasiswa PPA, serta laju error yang didapatkan dari pengujian ini sebesar 18,33%. Tingkat akurasi pada setiap jenis beasiswa dapat berbeda – beda tergantung pada jumlah pendaftar dan karakteristik dari beasiswa tersebut.

10. Penerapan Metode Algoritma C4.5 Untuk Prediksi Mahasiswa Non Aktif

(Irmayansyah & Lastrini, 2021). Data yang digunakan dalam penelitian ini adalah data berdasarkan kriteria yang digunakan dalam perhitungan, yaitu data mahasiswa tahun akademik 2016/2017 sebanyak 172. Metode yang digunakan adalah metode algoritma C4.5 dengan kriteria yang digunakan sebagai berikut : NPM, Nama, Tempat Lahir, Tanggal Lahir, Agama, Jenis Kelamin, Warga Negara. Data selection adalah Seleksi yang dilakukan adalah dengan menghapus field – field yang tidak digunakan, menghapus isi field yang kosong, dan menambahkan data yang memiliki kontribusi terhadap tujuan penelitian seperti data Indeks Prestasi Semester 7, Absensi semester 7, Sumber biaya, Penghasilan Per Tahun, Status Pembayaran dan field keterangan setiap mahasiswa. Dengan menerapkan pemodelan Algoritma C4.5 menjadi lebih efektif untuk proses pendataan ketidakaktifan mahasiswa. Telah dilakukan uji pengguna dengan menggunakan PSSUQ, dengan sesuai kategori diantaranya kategori Skor kepuasan secara keseluruhan (OVERALL) sebesar 87,78%, kegunaan sistem (SYSUSE) sebesar 89,28%, kualitas informasi (INFOQUAL) sebesar 87,24% dan kualitas antarmuka (INTERQUAL) sebesar 85,71%. Penelitian ini menghasilkan nilai akurasi dari sistem prediksi dengan menggunakan confusion matrix sebesar 81%.

Tabel 2. 1 Penelitian Rujukan

No	Peneliti	Judul	Jurnal	Kontribusi
1	Dian Ardiansyah, Walimah	ALGORITMA C4.5 UNTUK KLASIFIKASI CALON PESERTA LOMBA CERDAS CERMAT SISWA SMP DENGAN MENGGUNAKAN APLIKASI RAPID MINER	JurnalInkofar Vol.1 No. 2, Des 2018 ISSN: 2615-3645 (Print) / 2581-2920 (Online) http://politeknikmeta.ac.id/meta/ojs/index.php/inkofar/article/view/29/45 Citation from google scholar (Ardiansyah & Walimah, 2018)	Kontribusi dalam penelitian ini adalah : Mengklasifikasikan data calon peserta lomba sebagai sarana untuk menerapkan algoritma C4.5 dalam proses seleksi calon peserta lomba cerdas cermat siswa SMP. Kelemahan pada penelitian ini yaitu : Melakukan proses perhitungan menggunakan Aplikasi Rapid Miner.
	RAHMAD FAUZI	Klasifikasi Siswa Yang Akan Mengikuti Olimpiade Sains Nasional (OSN) Menggunakan Algoritma C4.5 (SMPN 2 Gresik)	http://eprints.umg.ac.id/2109/ Citation from google scholar : (Fauzi, 2018)	Kontribusi dalam penelitian ini adalah : konsep output yang diharapkan untuk merekomendasi siswa yang akan terpilih mengikuti lomba. Kelemahan penelitian ini adalah : Tidak dipublikasikannya langkah langkah perhitungan C4.5
3	Nadia Zulfa Rahma, Andik Setyono	Penerapan Algoritma C4.5 Dalam Memprediksi Kesiapan Siswa SMP IT PAPB Semarang Menghadapi UN	ISSN : (P)2087-7897 (O)2460-5344 Vol 8, No1 http://sisfotenika.stmikpontianak.ac.id/index.php/ST/article/view/177 (Rahma & Setyono, 2018)	Kontribusi dalam penelitian ini adalah : Metode klasifikasi dapat diterapkan untuk memprediksi kesiapan siswa SMP IT PAPB Semarang menghadapi UN. Kelemahan penelitian ini adalah : Karena pandemi, tidak ada lagi Tryout dan UN.

No	Peneliti	Judul	Jurnal	Kontribusi
4	Lastri Yuningsih, wan Rizal Setiawan, Asril Adi Sunarto	Rancangan Aplikasi Prediksi Kelulusan Siswa Menggunakan Algoritma C4.5	e-ISSN: 2685- 0877 p-ISSN:0216-3284 Vol 16, No 2 http://ojs.stmik-banjarbaru.ac.id/index.php/progresif/article/view/517 (Yuningsih, Setiawan, Sunarto, 2020)	Kontribusi dalam penelitian ini adalah : Dari hasil uji coba,didapatkan indeks prestasi gain tertinggi yaitu di semester 2 dengan nilai gain 0.31667. Kelemahan penelitian ini adalah : Walaupun telah menghasilkan rule yang dapat menjadi pedoman & dalam mengambil suatu keputusan, rule yang dihasilkan tersebut masih perlu pengujian faktual untuk mengetahui tingkat akurasi.
5	Yeni Angraini, Siti Fauziah, Jordi Lasmana Putra	ANALISIS KINERJA ALGORITMA C4.5 DAN NAÏVE BAYES DALAM MEMPREDIKSI KEBERHASILAN SEKOLAH MENGHADAPI UN	P-ISSN: 2685- 8223 E-ISSN: 2527- 4864 http://ejournal.nusamandiri.ac.id/index.php/jitk/article/view/1233/615 Citation from google scholar : (Angraini, Fauziah, & Putra, 2020)	Dibanding jurnal ke (3) Kontribusi dalam penelitian ini adalah Menghasilkan jawaban atas permasalahan dengan output yaitu terlihat perbedaan akurasi algoritma C4.5 dan Naive Bayes. Dengan dilakukan evaluasi secara confusion matrix terbukti bahwa pengujian yang dilakukan oleh Naïve Bayes memiliki nilai akurasi yang lebih tinggi dibanding C4.5. Nilai akurasi untuk model Naïve Bayes sebesar 95,50% dan nilai akurasi model algoritma C4.5 sebesar 78,50%. Diperoleh selisih akurasi sebesar 17%.
6	Castaka Agus Sugianto	Penerapan Teknik Data Mining Untuk Menentukan Hasil Seleksi Masuk SMAN 1 Gibeber Untuk Siswa Baru Menggunakan Decision Tree.	TEDC Vol.9 No.1 Jan 2015: 39-43 http://ejournal.poltectedc.ac.id/index.php/tedc/article/view/240 (Sugianto, 2019)	Kontribusi dalam penelitian ini adalah : Tingkat akurasi penentuan prediksi siswa yang diterima cukup tinggi yaitu mencapai sebesar 99.05%.Kelemahan dalam penelitian ini adalah :Atribut yang dipakai hanya nilai UN dan jumlah nilai UN.

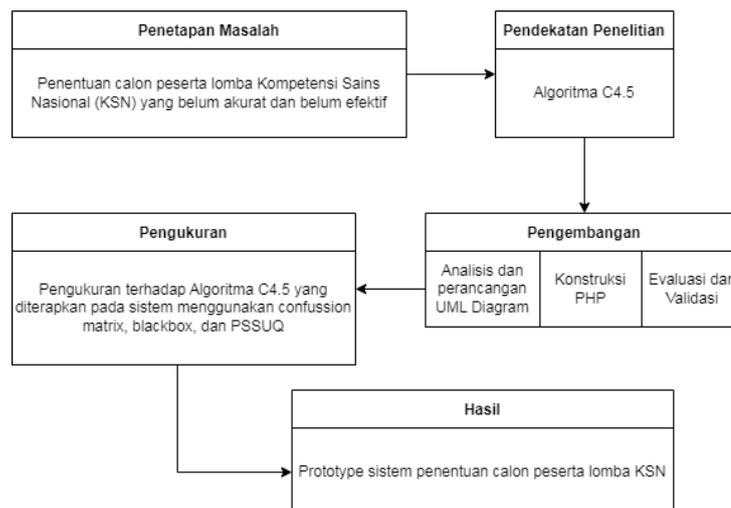
No	Peneliti	Judul	Jurnal	Kontribusi
7	Fardiansa h Putra	Penerapan Algoritma C4.5 dalam Menentukan Kandidat Siswa/Siswi SMK dalam Mengikuti Lomba Kompetensi Siswa.	https://repositori.usu.ac.id/handle/123456789/18422 Citation from google scholar : (Putra, 2016)	Kontribusi dalam penelitian ini adalah : Membuat suatu pola keputusan yang dijadikan landasan, dengan melakukan proses penambangan data pemilihan kandidat siswa pada Sekolah Nurul Amaliyah Tanjung Morawa untuk perhitungan dengan menggunakan rumus C4.5. Kelemahan dalam penelitian ini adalah : Metode decision tree yang digunakan sebaiknya mampu mengatasi keberadaan missing value. Hendaknya untuk pengembangan lebih lanjut, harap memperhatikan kelengkapan data faktor yang lain yang lebih baik agar hasil yang didapat semakin baik.
8	Hozairi Hozairi, Anwari Anwari, Syariful Alim	Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes.	P-ISSN: 2355- 2190/E-ISSN: 2615- 6539 NERO: https://nero.trunojoyo.ac.id/index.php/nero/article/view/237 (Hozairi, Anwari, & Alim, 2021)	Kontribusi dalam penelitian ini adalah : Mendeteksi sejak awal kondisi mahasiswa supaya kelulusannya tidak terlambat dan mempengaruhi nilai akreditasi Program Studi Teknik Informatika Universitas Islam Madura.
9	Ainur Rofik, Dedy Rahman Prehanto	Sistem Informasi Pendukung Keputusan Penerimaan dan Monitoring Beasiswa Dengan Algoritma C4.5 (Universitas Negeri Surabaya)	(JEISBI) Vol 1-1 ISSN 2774-3993 https://ejournal.unesa.ac.id/index.php/JEISBI/article/view/36799 (Rofik & Prehanto, 2020)	Kontribusi dalam penelitian ini adalah : Dari hasil pengujian yang dilakukan tingkat akurasinya 81,67% dan laju error 18,33 %. Presentase akurasi dan laju error didapat dari hasil proses seleksi penerimaan beasiswa menggunakan algoritma C4.5 dibandingkan dengan data learning yang telah ada.

No	Peneliti	Judul	Jurnal	Kontribusi
10	Irmayansyah, Erisya Lastrini	Penerapan Metode Algoritma C4.5 Untuk Prediksi Mahasiswa Non Aktif	Vol 11, No 2 (2021) P-ISSN 2087-3891 E-ISSN 2597-8918 http://www.teknois.stikombinaniaga.ac.id/index.php/JBS/article/view/119 (Irmayansyah & Lastrini, 2021)	Kontribusi dalam penelitian ini adalah : Uji pengguna dengan menggunakan PSSUQ, Skor kepuasan keseluruhan (OVERALL) sebesar 87,78%, kegunaan sistem (SYSUSE) sebesar 89,28%, kualitas informasi (INFOQUAL) sebesar 87,24% dan kualitas antarmuka (INTERQUAL) sebesar 85,71%. Nilai akurasi confusion matrix sebesar 81%.

Berdasarkan penelitian rujukan diatas, maka penelitian ini dilakukan dengan permasalahan yang serupa serta dengan metode penyelesaian yang sama yaitu Algoritma C4.5. Tetapi perbedaan penelitian kali ini dilakukan dengan menggunakan variabel yang berbeda dengan jurnal sebelumnya yaitu, NIS, Nama, Nilai Matematika kelas 7, Nilai IPA kelas 7, Nilai IPS kelas 7, Nilai Matematika kelas 8, Nilai IPA kelas 8, Nilai IPS kelas 8, dan Ekskul. Prototype sistem aplikasi yang berbasis website sebagai pengembangannya, lalu diuji kembali dengan blackbox.

E. Kerangka Pemikiran

Berikut adalah kerangka pemikiran untuk memecahkan masalah penelitian ini yang dapat dilihat pada Gambar 2.5 :



Gambar 2. 5 Kerangka Pemikiran

Kerangka pemikiran dapat dijelaskan sebagai berikut :

- (a) Penetapan masalah mencakup fenomena mengenai penentuan calon peserta lomba KSN, lalu dapat diidentifikasi masalah diantaranya yaitu belum akurat dalam penentuan calon peserta lomba KSN dan belum efektifnya proses dalam penentuan calon peserta lomba KSN;
- (b) Dari permasalahan tersebut, peneliti ingin menyelesaikan masalah dengan menggunakan pendekatan penelitian dengan menerapkan metode Algoritma C4.5;
- (c) Pengembangan yaitu tahap melakukan analisis dan perancangan gambar untuk UML diagram, kemudian menentukan bahasa pemrograman dengan konstruksi PHP dan melakukan evaluasi serta validasi;
- (d) Pengukuran yaitu menguji ketepatan hasil perhitungan algoritma C4.5 dengan menggunakan confusion matrix, pengujian sistem kepada ahli sistem menggunakan blackbox, dan uji kebergunaan kepada pengguna menggunakan PSSUQ;
- (e) Hasil yaitu sistem informasi menampilkan hasil penentuan calon peserta lomba KSN.

F. Hipotesis Penelitian

Berdasarkan permasalahan yang dihadapi yaitu belum akurat dan belum efektif pada saat melakukan proses penentuan calon peserta Kompetisi Sains Nasional (KSN) di Sekolah Menengah Pertama (SMP), maka perlu adanya suatu cara untuk mengatasi permasalahan tersebut. Dalam teori data mining ada berbagai macam metode yang dapat melakukan pengklasifikasian dalam memperkirakan kejadian di masa depan berdasarkan pengalaman di masa lampau, salah satunya adalah metode algoritma C4.5. Metode Algoritma C4.5 adalah salah satu algoritma yang dapat melakukan pengklasifikasian untuk memprediksi suatu peluang di masa depan berdasarkan dari pengalaman dimasa sebelumnya. Algoritma C4.5 sudah banyak dikenal dan digunakan untuk memprediksi nilai atribut bertipe diskret dari record yang baru di penelitian-penelitian sebelumnya dengan permasalahan yang serupa. Penelitian yang telah dilakukan oleh Rahmad Fauzi (2018) di dalam penelitiannya yang berjudul "KLASIFIKASI SISWA YANG AKAN MENGIKUTI LOMBA OLIMPIADE SAINS NASIONAL (OSN) MENGGUNAKAN ALGORITMA C4.5" atribut yang digunakan yaitu Matematika (MTK), Ilmu Pengetahuan Alam (IPA), Ilmu Pengetahuan Sosial (IPS), dan Bahasa Inggris. Berdasarkan hal tersebut, maka dapat ditetapkan hipotesis pada penelitian ini penerapan Algoritma C4.5 diduga dapat menentukan calon peserta Kompetisi Sains Nasional (KSN) di Sekolah Menengah Pertama (SMP).