

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Studi

Tinjauan studi merupakan analisis pustaka yang menggambarkan, menilai, serta menampilkan hasil dan gagasan utama dari riset sebelumnya terhadap sebuah topik.

2.1.1 Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Produktivitas Hasil Pertanian Bawang Merah

Dalam Salah satu sumber ekonomi bangsa adalah pertanian, meningkatkan produktivitas lahan pertanian dapat meningkatkan kesejahteraan petani. Bawang merah (*Allium cepal*) adalah sejenis tanaman yang menjadi bumbu berbagai masakan Asia Tenggara dan dunia. Resiko produksi bersifat risk increasing dan juga risk decreasing. Selain resiko produksi faktor lingkungan juga dapat timbul dan akan lebih sulit untuk mengontrolnya.

Salah satu metode yang menggunakan konsep probabilistik adalah naive bayes. Algoritma naive bayes adalah salah satu algoritma dalam teknik klasifikasi yang mudah diimplementasikan dan cepat prosesnya. Pada metode ini, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain. Tahap awal cara kerja dari proses perhitungan naive bayes adalah dengan melakukan pengambilan data training dari data petani. Adapun variabel penentu yang digunakan dalam mengklasifikasikan data petani yaitu jenis bibit, pengairan, penyakit, pengobatan dan pemupukan.

Tabel 2. 1 Hasil Pertanian Bawang Merah

No	Benih	Pupuk	Penyakit	Pengobatan	Pengairan	Hasil
1	Bagus	Cukup	Normal	Sangat Cukup	Cukup	Panen
2	Bagus	Kurang	Mboler	Sedang	Kurang	Gagal
3	Bagus	Kurang	Ulat	Sangat Cukup	Kurang	Gagal
4	Kurang Bagus	Cukup	Normal	Kurang	Cukup	Panen
5	Kurang Bagus	Cukup	Ulat	Kurang	Cukup	Gagal
6	Kurang Bagus	Kurang	Normal	Kurang	Kurang	Gagal
7	Bagus	Cukup	Ulat	Sedang	Cukup	Panen
8	Bagus	Cukup	Mboler	Sangat Cukup	Cukup	Panen
9	Bagus	Cukup	Mboler	Kurang	Cukup	Gagal
10	Bagus	Cukup	Ulat	Kurang	Cukup	Panen
11	Bagus	Cukup	Normal	Kurang	Kurang	Panen
12	Kurang Bagus	Cukup	Ulat	Sedang	Cukup	Panen
13	Kurang Bagus	Cukup	Mboler	Sedang	Kurang	Gagal
14	Kurang Bagus	Kurang	Ulat	Sangat Cukup	Cukup	Panen
15	Kurang Bagus	Kurang	Mboler	Sangat Cukup	Cukup	Panen

Menghitung jumlah probabilitas hasil dan atau gagal panen.

$P(Y=\text{Panen}) = 9/15 : 60\%$ “Jumlah data hasil panen pada data training dibagi dengan jumlah keseluruhan data”.

$P(Y=\text{Gagal}) = 6/15 : 40\%$ “Jumlah data hasil gagal panen pada data training dibagi dengan jumlah keseluruhan data”.

Pada tahap ini dilakukan penentuan data tasting untuk memperoleh hasil prediksi. Nilai atribut yang ditentukan seperti apa, kalikan semua nilai atribut untuk dapat menghasilkan hasil prediksi yang dicari. Semisal menghitung data tasting yaitu Benih (bagus) * Pupuk (cukup) * Penyakit (normal) * Pengobatan (sangat cukup) * Pengairan (cukup). Maka hasilnya sebagai berikut:

$$\text{Panen} : 56\% * 78\% * 33\% * 44\% * 89\% = 3,4\%$$

$$\text{Gagal} : 50\% * 50\% * 17\% * 17\% * 33\% = 0,1\%$$

Dari hasil tersebut menunjukkan presentase hasil panen 3,4% dan gagal 0,1%. Jadi hasil dari yang dicari adalah panen[8].

Adapun perbedaan antara penelitian di atas dengan penelitian yang penulis lakukan terletak pada objek penelitian, kriteria yang digunakan, serta pengklasifikasinya. Diharapkan hasilnya juga memiliki perbedaan.

2.1.2 Prediksi Jumlah Hasil Panen Sawit Menggunakan Algoritma Naive Bayes

Kelapa sawit merupakan salah satu tumbuhan penghasil minyak yang memiliki banyak kegunaan diantaranya sebagai bahan bakar, minyak industri, serta bahan baku minyak masak yang digunakan sehari-hari. Fenomena yang terjadi adalah sejak beberapa tahun belakangan, harga salah satu jenis minyak nabati ini cenderung stagnan dan menurun. Namun di sisi lain dapat menjadi waktu yang tepat untuk meningkatkan kembali harga CPO (Crude Palm Oil) salah satunya dengan melakukan peramalan jumlah produksi tanaman kelapa sawit untuk periode ke depan sehingga dapat menjadi acuan dalam menentukan kebijakan dan strategi bisnis, hal ini perlu agar setiap permintaan akan kebutuhan kelapa sawit tersedia.

Seperti yang tertera pada tabel, data yang digunakan adalah data numerik. Dalam langkah-langkah penyelesaian naive bayes, jika data yang digunakan merupakan data numerik maka langkah pertama yang harus dilakukan adalah Mencari Nilai Mean dan Standar Deviasi. Sebelum mencari nilai mean dan standar deviasi, harus dilakukan terlebih dahulu jumlah kelas meningkat dan menurun berdasarkan data training yang digunakan. Adapun

variabel penentu yang digunakan dalam mengklasifikasikan data hasil panen yaitu pokok awal, sensus ganoderma, pokok produktif dan produksi.

Tabel 2.2 Hasil Panen Sawit

No	Bulan	Pokok Awal	Sensus Ganoderma	Pokok Produktif	Produksi	Hasil
1	Januari	10754	4138	6616	453190	Menurun
2	Ferbruari	15443	2353	13090	446530	Menurun
3	Maret	18482	4186	14296	564350	Menurun
4	April	46715	5391	41324	564350	Meningkat
5	Mei	2315	481	2796	289510	Menurun
6	Juni	14452	5148	9304	649620	Menurun
7	Juli	41624	9753	31871	501010	Meningkat
8	Agustus	29529	4625	24904	594490	Meningkat
9	September	26751	3306	23445	357820	Menurun
10	Oktober	8766	4623	4143	375410	Menurun
11	November	10535	782	9753	404850	Menurun
12	Desember	23287	4716	18571	413680	Menurun

$$P(\text{Meningkat}|\text{Januari}) = P(\text{Meningkat}) * P(\text{Januari}|\text{Meningkat})$$

$$P(\text{Meningkat}|\text{Januari}) = 0,25 * 0,2437533$$

$$P(\text{Meningkat}|\text{Januari}) = 0,609383$$

$$P(\text{Menurun}|\text{Januari}) = P(\text{Menurun}) * P(\text{Januari}|\text{Menurun})$$

$$P(\text{Menurun}|\text{Januari}) = 0,75 * 0$$

$$P(\text{Menurun}|\text{Januari}) = 0$$

Dari hasil tersebut menunjukkan presentase hasil meningkat 0,609383 dan menurun 0. Jadi hasil dari prediksi adalah meningkat[9].

Adapun perbedaan antara penelitian di atas dengan penelitian yang penulis lakukan terletak pada objek penelitian, kriteria yang digunakan, serta pengklasifikasinya.

2.1.3 Analisa Data Pertanian Tanaman Pangan Untuk Memprediksi Hasil Panen Dengan Data Mining Algoritma C.45 (Studi Kasus : Dinas Tanaman Pangan dan Hortikultura Provinsi Sumut)

Data statistik Indonesia tahun 2015 menunjukkan Sumatera Utara merupakan provinsi penghasil padi terbesar ketiga setelah Pulau Jawa. Ditahun 2015, luas panen Sumut mencapai 71.756 hektare dengan produksi 857.553 ton. Sedangkan ditahun 2016, dengan luas panen yang lebih sedikit, yakni 69.960 hektare justru produksinya tetap di atasnya, yakni mencapai 907.908 ton. Dilihat dari produktivitasnya, tahun 2015 sebesar 119,51 ton per hektare sedangkan di tahun 2016 naik menjadi 129,78 ton per hektare. Namun lemahnya teknologi pendukung menjadi salah satu kendala peningkatan hasil panen, kurangnya sarana prediksi hasil panen mengakibatkan kurangnya informasi yang dibutuhkan untuk meningkatkan hasil panen.

Dalam metode algoritma C4.5, penentuan simpul awal dilakukan melalui perhitungan jumlah kasus. Jumlah kasus dengan hasil panen meningkat, stabil dan menurun untuk setiap atribut, entropy dari semua kasus dan gain dari setiap atribut. Berikut tabel pembagian kasus dari atribut sasaran tanam (luas, sedang, sempit), hasil panen (meningkat, stabil, menurun), curah hujan (tinggi, sedang rendah), hama (ya, tidak).

Tabel 2.3 Hasil Panen Padi

No	Sasaran Tanam (Ha)	Hasil Panen (Ton) 2015	Hasil Panen (Ton) 2016	Curah Hujan	Hama	Hasil
1	30	144289	150104	Tinggi	Tidak	Meningkat
2	89	484040	503547	Sedang	Tidak	Meningkat
3	106	584057	607595	Sedang	Tidak	Meningkat
4	99	580793	604199	Tinggi	Tidak	Meningkat
5	19	145416	151276	Tinggi	Tidak	Meningkat
6	30	106222	110503	Tinggi	Tidak	Meningkat

$$\text{Gain}(\text{Total Nilai Hasil Panen Priode April-Sept'16}) = -0.1501 - ((1.14 \times -0.3010) + (2.14 \times -0.3010) + (11.14 \times -0.1518)) = -0.1501 - (-0.0214) + (-0.0430) + (-0.1193) = -0.291$$

Dari hasil Tabel dapat diketahui bahwa atribut dengan gain tertinggi adalah hasil panen yaitu sebesar 0.291 dengan demikian, Nilai total hasil panen menjadi Node akar. Ada tiga atribut pada Nilai Akhir, yaitu 50. Dari tiga nilai atribut tersebut, ketiga atribut tersebut sudah mengklasifikasikan menjadi satu, yaitu meningkat[10].

Adapun perbedaan antara penelitian di atas dengan penelitian yang penulis lakukan terletak pada kriteria yang digunakan, serta pengklasifikasinya namun memiliki objek penelitian yang sama.

2.1.4 Sistem Pendukung Keputusan untuk Prediksi Produksi Cengkeh Menggunakan Metode Naïve Bayes

Cengkeh digunakan sebagai bahan baku dalam industri rokok, obat-obatan, makanan, minuman, dan kosmetik. Produksi cengkeh yang stabil dan efisien sangat penting dalam memenuhi permintaan pasar yang semakin meningkat. Namun, produksi cengkeh dipengaruhi oleh banyak faktor yang kompleks, seperti kondisi iklim, jenis tanah, pemupukan, pengairan, dan pengelolaan hama dan penyakit. Kekurangan informasi yang akurat dan prediksi yang dapat diandalkan seringkali mengakibatkan ketidakpastian dan risiko yang tinggi dalam produksi cengkeh.

Dalam konteks prediksi, metode Naive Bayes dapat digunakan untuk memprediksi berbagai variabel tergantung pada data input dan label yang tersedia. Misalnya, dalam prediksi produksi cengkeh, metode Naive Bayes dapat digunakan untuk memprediksi variabel seperti tingkat produksi berdasarkan pada fitur-fitur seperti luas lahan, cuaca, jumlah produksi tahun sebelumnya, dan jumlah pekerja.

Tabel 2.4 Hasil Panen Cengkeh

No	Produksi	Cuaca	Luas Lahan	Pekerja	Hasil
1	Kurang	Buruk	Sedikit	Banyak	Turun
2	Banyak	Buruk	Banyak	Banyak	Naik
3	Banyak	Buruk	Sedikit	Banyak	Naik
4	Banyak	Baik	Banyak	Sedikit	Naik
5	Kurang	Baik	Banyak	Sedikit	Turun
6	Banyak	Baik	Banyak	Banyak	Naik

$$P(\text{keterangan naik}) = 12/20 = 0.60$$

$$P(\text{keterangan turun}) = 8/20 = 0.40$$

$$P(\text{naik}) = (0.60) * (1) * (0.5) * (0.75) * (0.75) = 0.16875$$

$$P(\text{turun}) = (0.40) * (1) * (0.5) * (0.5) * (0.5) = 0.05$$

Dari hasil tersebut menunjukkan presentase hasil naik 0.16875 dan turun 0.05. Jadi hasil dari prediksi adalah naik[11].

Adapun perbedaan antara penelitian di atas dengan penelitian yang penulis lakukan terletak pada objek penelitian, kriteria yang digunakan, serta pengklasifikasinya.

2.2 Tinjauan Pustaka

2.2.1 Data Mining

Data mining adalah suatu pencarian dan analisa pada suatu koleksi data sehingga ditemukan suatu pola yang menarik dengan tujuan mengekstrak informasi dan pengetahuan yang akurat dan potensial,serta dapat dipahami dan berguna bagi pengambilan keputusan[12].

2.2.2 Prediksi

Prediksi adalah suatu proses perkiraan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi[13].

2.2.3 Padi

Padi (*Oryza sativa* L.) merupakan tanaman utama dunia, karena menjadi makanan pokok di beberapa negara termasuk Indonesia dan berperan penting dalam perekonomian pada sebagian besar negara-negara berkembang. Komoditas tanaman padi dapat dibudidayakan pada dua jenis lahan yaitu sawah dan ladang[14].

2.2.4 Panen

Panen adalah kegiatan mengambil hasil dari tanaman setelah mencapai kemasakan optimal atau mempunyai potensi maksimal jika akan diolah menjadi bahan baku untuk industri atau langsung konsumsi. Untuk

tanaman padi dimulai dari proses persemaian sampai dengan padi dapat di panen dibutuhkan waktu 7 sampai dengan 8 bulan. faktor - faktor yang mempengaruhi hasil panen tanaman padi antara lain seperti luas tanam, faktor iklim, hama dan pemupukan[15]. Nilai indeks ambang batas (*benchmark yield*) hasil panen padi di tentukan berdasarkan data panen rata rata historis berdasarkan data balai penyuluh pertanian[16].

2.2.5 Naïve Bayes

Naïve bayes adalah salah satu algoritma yang populer digunakan untuk keperluan data mining karena kemudahannya serta pemrosesan yang cepat, mudah di implementasikan dengan strukturnya yang cukup sederhana dan tingkat efektifitas yang tinggi[17]. Adapun rumus dalam naive bayes adalah sebagai berikut[18].

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Keterangan :

X merupakan data dengan class yang belum diketahui.

H : hipotesis data X merupakan suatu class spesifik.

P (H|X) : Probabilitas hipotesis H berdasarkan kondisi.

X P(H) : Probabilitas hipotesis.

H P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis.

H P(X): Probabilitas X

Naïve bayes merupakan algoritma yang dapat mengklasifikasi suatu variabel tertentu dengan menggunakan metode probabilitas dan statistic[19].

2.2.6 Website

Website merupakan kumpulan halaman-halaman yang digunakan untuk menampilkan informasi berupa teks, gambar diam atau gerak, animasi, suara, dan gabungan dari semuanya, baik yang bersifat statis maupun dinamis, juga membentuk satu rangkaian bangunan yang berkaitan[20]. Website adalah aplikasi yang dapat dijalankan dengan menggunakan web browser, saat ini hampir semua gawai dapat menjalankan web browser yang menyebabkan website dapat dibuka di hampir semua gawai yang ada[21].

2.2.7 UML

UML adalah salah satu standar bahasa yang banyak digunakan di dunia industri untuk mendefinisikan requirement, membuat analisis dan desain, serta menggambarkan arsitektur dalam pemrograman berorientasi objek[22].

a. Diagram Use Case

Use case diagram merupakan pemodelan untuk kelakuan sistem informasi yang akan dibuat. Use case bekerja dengan mendeskripsikan tipikal interaksi antara user sebuah sistem dengan sistemnya sendiri melalui sebuah cerita bagaimana sistem itu dipakai[22].

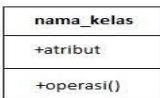
Tabel 2.5 Diagram Use Case

Simbol	Keterangan
	Aktor : Mewakili Peran Orang, sistem atau alat saat berkomunikasi dengan use case diagram.
	<i>Use Case</i> : Abstraksi dan Interaksiantara sistem dan actor.
	Asosiasi : Komunikasi antara aktordan use case yang berpartisipasi.
	Generalisasi : Use case merupakangeneralisasi yang lain.

b. Class Diagram

Class diagram adalah sebuah spesifikasi yang jika diinstansiasi akan menghasilkan sebuah objek dan merupakan inti dari pengembangan dan desain berorientasi objek. Class diagram menggambarkan struktur dan deskripsi class, package dan object beserta hubungannya satu sama lain seperti containment, pewarisan, asosiasi dan lain-lain. Atribut dan metode dapat memiliki salah satu sifat, yaitu : private, protected, public dan package[22].

Tabel 2.6 Class Diagram

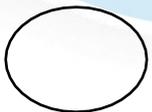
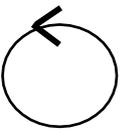
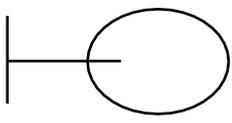
Simbol	Nama	Deskripsi
	<i>Class</i>	Kelas pada struktur sistem.

	Asosiasi / <i>Association</i>	Relasi antar kelas dengan makna umum, asosiasi biasanya juga disertai dengan <i>multiplicity</i> .
	Asosiasi berarah / <i>Directed association</i>	Relasi antar kelas dengan makna kelas yang satu digunakan oleh kelas yang lain, asosiasi biasanya juga disertai dengan <i>multiplicity</i> .
	Generalisasi	Relasi antar generalisasi – generalisasi (umum khusus).
	Kebergantungan / <i>dependency</i>	Relasi antar kelas dengan makna kebergantungan antar kelas.

c. Sequence Diagram

Sequence diagram mendokumentasikan komunikasi/interaksi antar-class. Diagram ini menunjukkan sejumlah objek dan message (pesan) yang diletakkan di antara objek-objek di dalam use case[22].

Tabel 2.7 Sequence Diagram

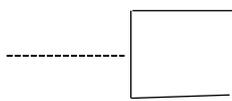
Simbol	Keterangan
	Aktor : Menggambarkan orang yang sedang berinteraksi dengan sistem.
	<i>Entity Class</i> : Menggambarkan hubungan yang terjadi.
	<i>Control Class</i> : Penghubung antaraboundary dan tabel
	Boundary Class, berisi kumpulan kelas yang menjadi interface atau interaksi satu atau lebih aktor dengan sistem

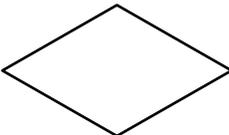
	A Fokus Of Control dan Life Line: Mulai dan berakhirnya message.
	A Message : Menggambarkan Pengiriman Pesan

2.2.8 Flowchart

Flowchart adalah cara untuk menjelaskan tahap-tahap pemecahan masalah dengan merepresentasikan simbol-simbol tertentu yang mudah dipahami, mudah digunakan dan standar. Tujuan penggunaan flowchart adalah untuk menggambarkan suatu tahapan penyelesaian masalah secara sederhana, teratur dan rapi dengan menggunakan simbol-simbol yang standar yang dapat di mengerti oleh programmer[23].

Tabel 2.8 Flowchart

Simbol	Keterangan
	Input/Output : Simbol yang menyatakan proses input dan output
	Proses : Simbol yang menyatakan suatu proses yang dilakukan komputer
	Penghubung : Simbol untuk keluar – masuk atau penyambungan proses yang sama
	Anak panah : Mempresentasikan operasi
	Penjelasan : Digunakan untuk komentar tambahan

	Keputusan : Untuk menunjukan dua kondisitertentu contohnya ya atau tidak
	<i>Preparation</i> : Memberikan Harga Awal
	Terminal Points : Awal atau Akhir dari <i>flowchart</i>
	Dokument : I/O dalam formt dicetak

2.2.9 Visual Studio Code

Visual Studio Code adalah Software yang sangat ringan, namun kuat editor kode sumbernya yang berjalan dari desktop. Visual Studio Code digunakan untuk pembuatan kode-kode program dibutuhkan sebuah aplikasi yang mumpuni. Visual Studio Code dapat digunakan untuk berbagai bahasa pemrograman seperti JavaScript, HTML, CSS, PHP, Python, C++, dan masih banyak lagi. Visual Studio Code bekerja pada berbagai sistem operasi seperti Windows, macOS, dan Linux[24].



Gambar 2.1 Visual Studio Code

2.2.10 Laragon

Laragon adalah perangkat lunak bebas yang mendukung banyak system operasi, berfungsi sebagai server diri sendiri / localhost. Laragon

ialah perangkat lunak bebas yang di dalamnya terdapat banyak sistem operasi sebagai localhost atau server mandiri. Laragon menyediakan banyak layanan, peralatan, dan fitur yang terdiri dari Apache, PHP Server, Phpmysql, Mysql, Memcached, Redis, Composer, Xdebug, Cmdre dan Laravel[25].

2.2.11 Mysql

MySQL adalah salah satu jenis database server yang sangat terkenal. MySQL menggunakan bahasa SQL untuk mengakses database nya. MySQL adalah sebuah database manajemen system (DBMS) populer yang memiliki fungsi sebagai relational database manajemen system (RDBMS)[26].



Gambar 2.2 Mysql

2.2.12 PHP

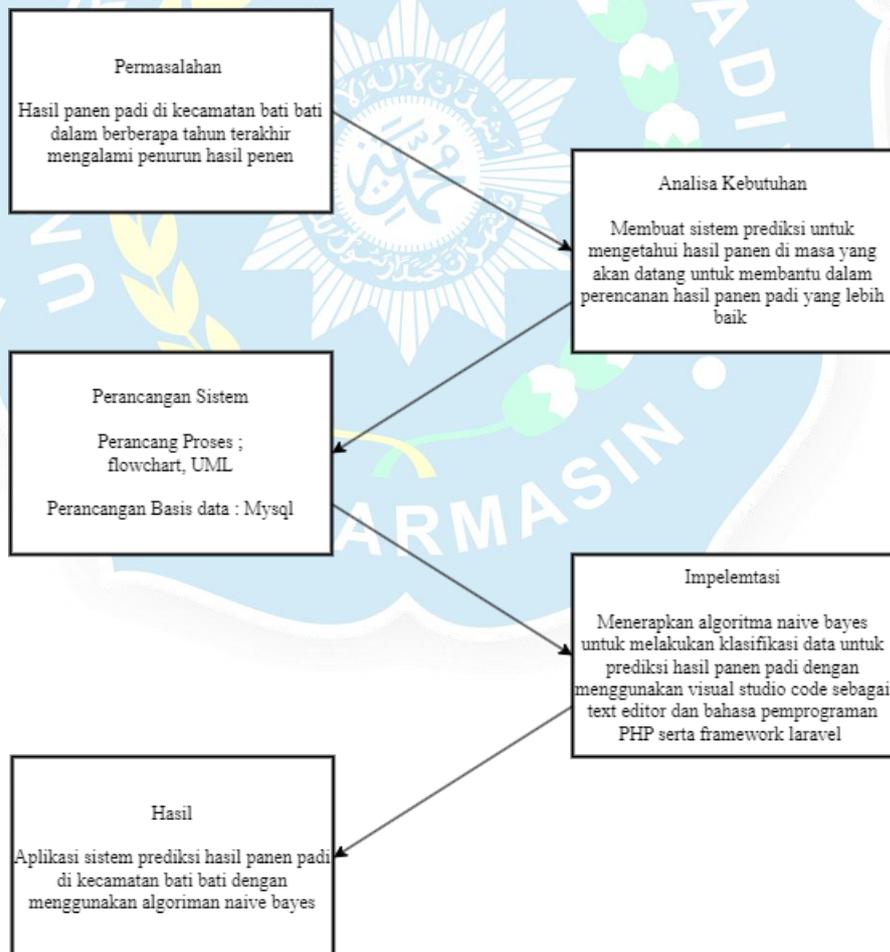
PHP yaitu sigkatan dari (Hypertext Preprocessor) merupakan bahasa pemrograman berbasis web yang memiliki kemampuan untuk memproses dan mengolah data secara dinamis . PHP adalah server-side embedded script language artinya, semua sintaks dan perintah program yang anda tulis akan sepenuhnya dijalankan oleh server, tetapi dapat disertakan pada halaman HTML biasa[27].

2.2.13 Laravel

Laravel adalah salah satu php framework paling populer dan paling banyak digunakan untuk membuat website saat ini. Dengan menggunakan laravel proses pengembangan aplikasi menjadi lebih cepat dan powerfull[28].

2.3 Kerangka Pemikiran

Kerangka pemikiran adalah narasi (uraian) atau pernyataan (proposisi) tentang kerangka konsep pemecahan masalah yang telah diidentifikasi atau dirumuskan. Kerangka pemikiran dapat dilihat pada gambar ini[29].



Gambar 2.3 Kerangka Pemikiran