

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Penelitian Terkait**

Penelitian yang dilakukan oleh Nugraha dkk [1] yang berjudul “*Adaptive E-learning system* Berbasis VARK *learning style* dengan klasifikasi materi pembelajaran menggunakan KNN (*K-Nearest Neighbor*)” pada tahun 2022. Hasil dari penelitian yang melibatkan 42 Mahasiswa X OTKP 3 ini, hasil dari kecenderungan belajar mahasiswa didominasi dengan gaya belajar aural sebanyak 38% dan *kinesthetic* sebanyak 50%, adapun mahasiswanya sebanyak 3% adalah mahasiswa dengan gaya belajar *visual dan read/write*. Evaluasi Model dengan menggunakan *confussion matrix* dengan dataset sebanyak 60, dengan pembagian data *training* sebanyak 48 dan data testing sebanyak 12, mendapatkan hasil yang baik, yaitu *accuracy* sebesar 91%, *pecission* sebesar 93% dan *recall* sebesar 91%, sedangkan untuk pengujian penggunaan sistem oleh responden menggunakan UAT mendapatkan presentase sebesar 81% .

Penelitian yang dilakukan oleh Dariyanti dkk [2] yang berjudul “Hubungan gaya belajar VAK (Visual, Auditori, Dan Kinestetik) dengan prestasi akademik mahasiswa Fakultas Kedokteran” pada tahun 2021. Hasil penelitian ini dari 156 responden, gaya belajar kinestetik sebanyak 68 orang (43,58%), gaya belajar auditori:55 orang (35,25%), gaya belajar *read-write*:22 orang (14,10%) dan gaya belajar visual:11 orang (7,05%). Hasil uji analisis bivariat diperoleh hasil  $r=0.829$  dan nilai  $p=0,829$  ( $p>0,05$ ), artinya satu gaya belajar tidak dapat membentuk prestasi akademik yang maksimal, tetapi bagaimana seseorang dapat 6

menggunakan gaya belajar yang diketahui sesuai dengan metode atau jenis pembelajaran yang dijalani [4].

Penelitian tentang klasifikasi gaya belajar telah dilakukan oleh beberapa peneliti sebelumnya. Diantaranya adalah penelitian yang dilakukan Hanun dkk[8]. Penelitian ini menggunakan 11 kriteria dalam penentuan gaya belajar siswa menggunakan metode *naive bayes*. Selain metode *naive bayes*, klasifikasi gaya belajar menggunakan metode yang lain juga dilakukan oleh beberapa peneliti terdahulu, seperti yang dilakukan Widya dkk[9]. Klasifikasi dilakukan menggunakan bantuan *tools* WEKA dan menghasilkan 23 *rule* dari bentuk *decision tree*. Penelitian ini menghasilkan rekomendasi penentuan gaya belajar yang baik. Penelitian lain yang sejenis juga pernah dilakukan oleh Agung dkk[10]. Penelitian ini menghasilkan tingkat akurasi 88% dengan menggunakan algoritma *k-nearest neighbour*. Berdasarkan penelitian terdahulu yang sejenis, maka peneliti menggunakan metode yang berbeda dari penelitian yang dilakukan Putu dkk [11] yaitu metode *naïve bayes classifier*.

Penelitian terkait pengelompokan sudah diadakan berbagai peneliti sebelumnya. Diantaranya penelitian oleh Sifiana dwi[12]. Penelitian ini mengelompokan, membandingkan, menganalisa kelompok tipe belajar siswa dan menarik analisis untuk membuat keputusan gaya belajar yang tepat untuk tiap kelas siswa. Dalam penelitian tersebut dua pendekatan algoritma yang digunakan yaitu, *k-means* dan FCM. Penelitian serupa juga pernah dilakukan oleh Nugrah dkk[4]. Hasil pengolahan *k-means* hanya membutuhkan waktu rata-rata 1 detik sedangkan pengolahan data pada *k-medoids* membutuhkan waktu rata-rata 1 menit

30 detik. Hasil penelitian tersebut menunjukkan bahwa nilai akurasi lebih besar yaitu 80% dan pada saat proses pengolahan data lebih efektif dan efisien menggunakan metode *k-means*. Dari penelitian sebelumnya yang sejenis, maka peneliti memakai *k-means clustering* karena algoritma ini mudah diimplementasikan dan dijalankan.

## **2.2 Landasan Teori**

### **2.2.1 Gaya Belajar**

Gaya belajar merupakan modalitas dalam kegiatan pembelajaran, dengan adanya sebuah gaya belajar, setiap individu dapat memahami kekurangan dan kelebihan saat belajar guna mendapatkan hasil belajar yang maksimal[5][6][7]. Terutama dalam pembelajaran bahasa, gaya belajar yang tepat akan membuat siswa lebih cepat memahami materi yang ada. Belajar merupakan poin utama dalam gaya belajar yang memiliki pengertian yaitu sebuah proses di mana pemahaman akan sesuatu yang merupakan akibat adanya interaksi stimulus dan *respons*, sehingga dapat menunjukkan perilaku dari hasil pembelajaran tersebut.

Murid bergaya belajar visual berkecenderungan belajar menggunakan indera penglihatan dibanding pendengaran[6]. Seseorang bergaya belajar visual belajar menitik beratkan pada ketajaman penglihatan sehingga upaya yang dapat dilakukan untuk menumbuhkan minat siswa pada saat proses pembelajaran berlangsung yaitu guru mampu menyajikan suatu informasi melalui gambar, video ataupun hal-hal lainnya secara tertulis. Karakteristik murid bergaya belajar visual, terdiri atas sangat rapi serta teratur, mementingkan penampilan, mengingat dengan gambar, sangat suka membaca, berbicara cepat, lebih suka seni daripada musik,

selalu mencermati gerak bibir pada saat berbicara dengannya, cenderung suka peragaan ketimbang penjelasan secara lisan, teliti, pengeja yang baik.

Mahasiswa bergaya belajar auditori paling mudah menangkap informasi yang mengandalkan pendengarannya. Seseorang dengan bergaya belajar auditori ini menitik beratkan pendengaran sebagai media guna menelaah informasi maupun pengetahuan[8]. Siswa dengan gaya belajar auditori biasanya sulit untuk menyerap pembelajaran secara langsung dalam bentuk tulisan, gambar dan lain sebagainya, sehingga metode ceramah ataupun memberikan informasi melalui audio adalah upaya yang dapat diimplementasikan pendidik selama kegiatan belajar berlangsung. Karakteristik murid bergaya belajar ciri-ciri auditori yaitu tidak suka kebisingan, pembicara yang fasih, suka belajar kelompok, menyukai musik, lebih mudah paham dengan mendengarkan, senang membaca dengan suara lantang, baik dalam berbicara maupun menyampaikan cerita, sangat bijak selama berbicara, serta belajar dengan mendengar maupun mengingat.

Mahasiswa yang memiliki gaya belajar kinestik mengharuskan individu analisis gaya belajar VAK pada pembelajaran daring terhadap minat belajar siswa mendapatkan informasi harus dengan terlibat langsung, sehingga belajar dengan cara praktek adalah upaya yang dapat diimplementasikan pendidik selama kegiatan belajar mengajar[8]. Ciri-ciri gaya belajar kinestik yaitu berbicara dengan pelan, banyak menggunakan isyarat tubuh, sangat suka berolahraga, mudah bosan [8]. Gaya belajar seseorang sangat mudah berubah tergantung dengan kebiasaan, untuk itu bukan hal yang mustahil jika gaya belajar siswa akan terpengaruh oleh cara mengajar seorang guru. Siswa dengan gaya belajar visual

bisa saja berubah menjadi auditori ataupun kinestetik, begitu pula sebaliknya siswa dengan gaya belajar kinestetik bisa saja berubah menjadi visual ataupun auditori.

### **2.2.2 Gaya Belajar VAK**

Model pembelajaran VAK adalah model yang menganggap bahwa pembelajaran akan efektif dengan memperhatikan ketiga hal yaitu *visual* (melihat), *auditory* (mendengar), dan *kinesthetic* (bergerak)[11][12], dengan kata lain memanfaatkan potensi siswa yang telah dimilikinya dengan melatih dan mengembangkannya. Pembelajaran dengan model ini mementingkan pengalaman belajar secara langsung dan menyenangkan bagi siswa. Pengalaman belajar secara langsung dengan cara belajar dengan mengingat (*visual*), belajar dengan mendengar (*auditory*), dan belajar dengan gerak dan emosi (*kinesthetic*).

Menurut Herdian, model pembelajaran VAK merupakan suatu model pembelajaran yang menganggap pelajaran akan efektif dengan memerhatikan ketiga hal tersebut (*visual, auditory, kinesthetic*) dan dapat diartikan bahwa pembelajaran dilaksanakan dengan memanfaatkan potensi siswa yang telah dimilikinya dengan melatih dan mengembangkannya[13][14].

### **2.2.3 Data Mining**

Data *mining* adalah pembelajaran mesin, pengenalan pola, *database*, statistik, dan teknik visualisasi yang digunakan untuk memecahkan masalah penggalian informasi dari repositori *database* besar[9]. Data *mining* sangat penting dalam proses penggalian data secara manual dari kumpulan data berupa pengetahuan yang tidak diketahui[9][10].

Berdasarkan pengertian data *mining* tersebut dapat disimpulkan bahwa data *mining* adalah suatu proses pencarian data secara otomatis dapat mendapatkan sebuah model dari *database* yang besar.

Adapun data yang tidak terlepas dari data mining yaitu data *warehouse* adalah kumpulan desain basis data terintegrasi dan berorientasi subjek untuk mendukung fungsi DSS, di mana setiap unit data adalah *non-volatile* dan relevan dengan waktu tertentu repositori fisik di mana data relasional diorganisasikan secara khusus untuk menyediakan data yang dibersihkan perusahaan dalam format standar data.

*Bussines Intelegen* adalah istilah umum yang menggabungkan arsitektur, alat, basis data, alat analisis, aplikasi, dan metodologi. Seperti DSS, BI ekspresi bebas konten, berarti itu hal yang berbeda untuk orang yang berbeda. Tujuan utama BI adalah untuk memungkinkan akses mudah ke data (dan model) untuk memberikan manajer bisnis dengan kemampuan untuk melakukan analisis. BI membantu mengubah data, menjadi informasi (dan pengetahuan), menjadi keputusan dan akhirnya bertindak.

Data *mining* menurut Suyanto adalah gabungan sejumlah disiplin ilmu komputer yang mendefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari *artificial intelligence*, *machine learning*, *statistics*, dan *database systems* [3].

Berdasarkan pengertian data *mining* menurut para ahli yang disebutkan maka dapat disimpulkan bahwa data *mining* adalah suatu proses pencarian data secara otomatis dapat mendapatkan sebuah model dari *database* yang besar.

#### 2.2.4 Tahapan data *mining*

Data *mining* merupakan metode untuk menemukan informasi tersembunyi dalam *database* dan bagian dari proses KDD untuk menemukan informasi dan pola yang berguna dalam data[4][5].

Secara umum, data *mining* dibagi menjadi dua kategori utama yaitu[6] :

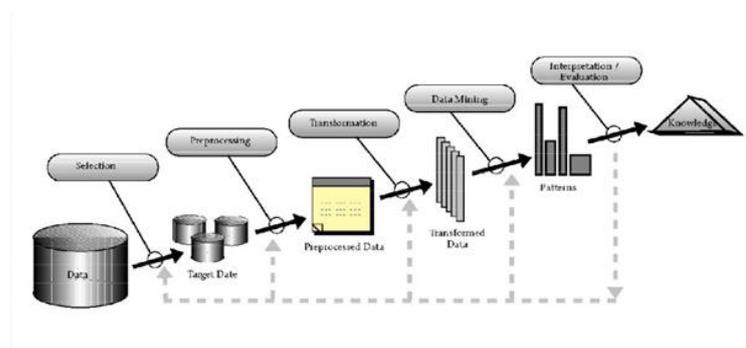
##### 1. Prediktif

Proses untuk menemukan karakteristik penting dari data dalam satu basis data. Teknik data mining yang termasuk *descriptive mining* adalah *clustering*, *asosiation*, dan *sequential mining*.

##### 2. Deskriptif

Proses untuk menemukan pola dari data dengan menggunakan beberapa variabel lain di masa depan. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi. Secara sederhana data mining biasa dikatakan sebagai proses penyaring atau “menambang” pengetahuan dari sejumlah data yang besar.

Secara keseluruhan, proses KDD dapat digambarkan sebagai berikut[7]:



Gambar 2.1 Proses KDD data mining [7]

1. *Data Selection* (seleksi data)

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data *mining*.

2. *Pre-processing*

Sebelum proses data *mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang konsisten, dan memperbaiki kesalahan pada data.

3. *Transformation* (pengubah data)

*Transformation* adalah mengubah data kedalam bentuk yang sesuai untuk ditambang, proses ini sering disebut transformasi data. Terdapat juga teknik dalam penambangan data yaitu ;

1. *Predictive modeling*, terdapat dua teknik yaitu *classification* dan *value prediction*.
2. *Database Segmentation*, melakukan partisi *database* menjadi sejumlah segmen, *cluster*, atau *record* yang sama.
3. *Link analysis*, sebuah teknik untuk membuat hubungan antara *record* yang individu atau sekumpulan *record* dalam *database*.
4. *Deviation detection*, sebuah teknik untuk mengidentifikasi *outlier* yang mengekspresikan sebuah deviasi dari ekspektasi yang sudah diketahui sebelumnya.

5. *Nearest Neighbour*, yaitu teknik yang memprediksi pengelompokan, teknik ini sendiri merupakan teknik yang tertua yang digunakan dalam *data mining*. *clustering*, merupakan teknik untuk mengklasifikasikan data berdasarkan kriteria masing-masing data, sehingga dalam suatu *cluster* merupakan data dengan karakteristik yang hampir sama. Analisis *clustering* adalah sebuah teknik dari analisis *multivariable* yang digunakan untuk mengelompokkan objek (*variable* atau data) sehingga dapat menghasilkan suatu informasi untuk membantu pelaksanaan pengujian terhadap objek dan pada akhirnya dapat menyajikan suatu hipotesis berdasarkan relasi yang terjadi. Tujuan dari teknik ini adalah untuk melakukan pengelompokan berdasarkan kriteria tertentu sehingga objek-objek tersebut mempunyai variasi di dalam *cluster* relatif kecil dibandingkan variasi antar cluster, untuk mengukur tingkat kesamaan tersebut digunakan suatu rumus untuk perhitungan jarak, rumus yang sering digunakan adalah rumus jarak *euclidean*.
6. *Decision tree* merupakan teknik generasi selanjutnya, dimana teknik ini adalah sebuah model prediktif yang dapat digambarkan seperti pohon. Setiap node yang terdapat dalam struktur pohon tersebut mewakili sebuah pertanyaan yang digunakan untuk menggolongkan data.

#### 4. Data Mining (Penambangan Data)

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

#### 5. Intepretasi/Evaluasi

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Seperti menggunakan visualisasi atau tampilan yang dapat menjelaskan luaran sistem..

### **2.2.5 Klasifikasi**

Klasifikasi merupakan salah satu tahap penting dalam data mining, klasifikasi adalah pengelompokan data atau objek baru ke dalam kelas atau label berdasarkan atribut-atribut tertentu [4]. Teknik dari klasifikasi adalah dengan melihat variabel dari kelompok data yang sudah ada. Klasifikasi bertujuan untuk memprediksi kelas dari suatu objek yang tidak diketahui sebelumnya. Klasifikasi terdiri dari tiga tahap, yaitu pembangunan model, penerapan model, dan evaluasi. Pembangunan model adalah membangun model menggunakan data latih yang telah memiliki atribut dan kelas, kemudian data-data tersebut diterapkan untuk menentukan kelas dari data atau objek yang baru, setelah itu data dievaluasi untuk melihat tingkat akurasi dari pembangunan dan penerapan model terhadap data baru [5]. Proses klasifikasi terdiri dari dua fase, yaitu fase *training* dan fase *testing*. Fase *training* adalah fase di mana data digunakan untuk membangun sebuah model sedangkan fase *testing* adalah pengujian model yang telah dibuat dengan data lainnya untuk mengetahui akurasi dari model tersebut.

### **2.2.6 Algoritma K-Nearest Neighbor**

*K-Nearest Neighbor* (KNN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari label *class* pada K-NN[4]. Algoritma *K-Nearest Neighbor* (KNN) merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan (K) tetangga

terdekatnya[3]. KNN termasuk algoritma *supervised learning*, yang mana hasil dari *query instance* baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Kelas yang paling banyak muncul, yang akan menjadi kelas hasil klasifikasi[3].

Tujuan dari algoritma K-NN adalah mengklasifikasikan objek baru berdasarkan atribut dan *training data*[4]. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak *euclidian*[11].

*Euclidean distance* mengukur kedekatan antara dua buah objek yang digambarkan sebagai garis lurus atau secara langsung [11]. Metode pengukuran ini cocok apabila diimplementasikan terhadap data yang memiliki nilai atribut bersifat *numerical*, khususnya dengan atribut kontinu[11]. *Euclidean distance* memiliki rumus persamaan sebagai berikut:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Keterangan :

d = Jarak

x1 = Koordinat *latitude* 1

x2 = Koordinat *latitude* 2

y1 = Koordinat *longitude* 1

y2 = Koordinat *longitude* 2

Langkah-langkah perhitungan *K- Nearest Neighbor* (KNN) sebagai berikut:

1. Menentukan parameter K (jumlah tetangga paling dekat).

2. Menghitung kuadrat jarak *eucliden* objek terhadap data *training* yang diberikan
3. Selanjutnya mengurutkan hasil no 2 secara *ascending* (berurutan dari nilai tinggi kerendah)
4. Mengumpulkan kategori *y* (klasifikasi *nearest neighbor* berdasarkan nilai (*k*))

Kelebihan dari algoritma KNN yaitu :

1. Mudah diterapkan

Mengingat kesederhanaan dan akurasi algoritma, KNN merupakan salah satu pengklasifikasi pertama yang sebaiknya dipelajari oleh data *scientist* pemula.

2. Mudah beradaptasi

Saat sampel *training* baru ditambahkan, algoritma KNN menyesuaikan untuk ikut memperhitungkan data baru karena semua data pelatihan disimpan ke dalam memori.

3. Memiliki sedikit *hyperparameter*

KNN hanya membutuhkan nilai *k* dan metrik jarak, yang relatif lebih sedikit jika dibandingkan dengan algoritma *machine learning* lainnya.

Kekurangan dari algoritma KNN yaitu :

1. Tidak berfungsi dengan baik pada dataset berukuran besar

*Dataset* berukuran besar, *cost* untuk menghitung jarak antara titik baru dan setiap titik yang ada sangat besar dan cenderung menurunkan kinerja algoritma.

2. Kurang cocok untuk dimensi tinggi

Algoritma KNN tidak bekerja dengan baik pada data berdimensi tinggi karena dengan jumlah dimensi yang besar, menjadi sulit bagi algoritma untuk menghitung jarak di setiap dimensi.

### 3. Perlu penskalaan fitur

Kita perlu melakukan penskalaan fitur (standarisasi dan normalisasi) sebelum menerapkan algoritma KNN ke kumpulan data apa pun. Jika kita tidak melakukannya, KNN dapat menghasilkan prediksi yang salah.

### 4. Sensitif terhadap *noise* data, *missing values* dan *outliers*

KNN sensitif terhadap *noise* dalam *dataset*, kita perlu secara manual memasukkan nilai yang hilang dan menghapus *outlier*.

## 2.2.7 Confusion Matrix

*Confusion Matrix* adalah alat evaluasi visual yang digunakan dalam pembelajaran mesin. Kolom matriks kebingungan mewakili hasil kelas prediksi, dan baris mewakili hasil kelas yang sebenarnya. Ini menyebutkan semua kemungkinan kasus masalah klasifikasi. Mengambil masalah klasifikasi biner sebagai contoh, dimensi dari matriks kebingungan adalah  $2 \times 2$ . Serangkaian ukuran kinerja algoritma dapat didefinisikan menggunakan matriks kebingungan, seperti tingkat pemeriksaan wilayah positif dan kelas *negative*. [11]

*Confusion matrix* adalah pengukuran performa untuk masalah klasifikasi *machine learning* dimana keluaran dapat berupa dua kelas atau lebih. *Confusion matrix* adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai ctual. Ada empat istilah yang merupakan representasi hasil

proses klasifikasi pada *confusion matrix* yaitu *true positif*, *true negatif*, *nalse positif*, dan *false negatif*.

Tabel 2.1 contoh *Confusion Matriks*

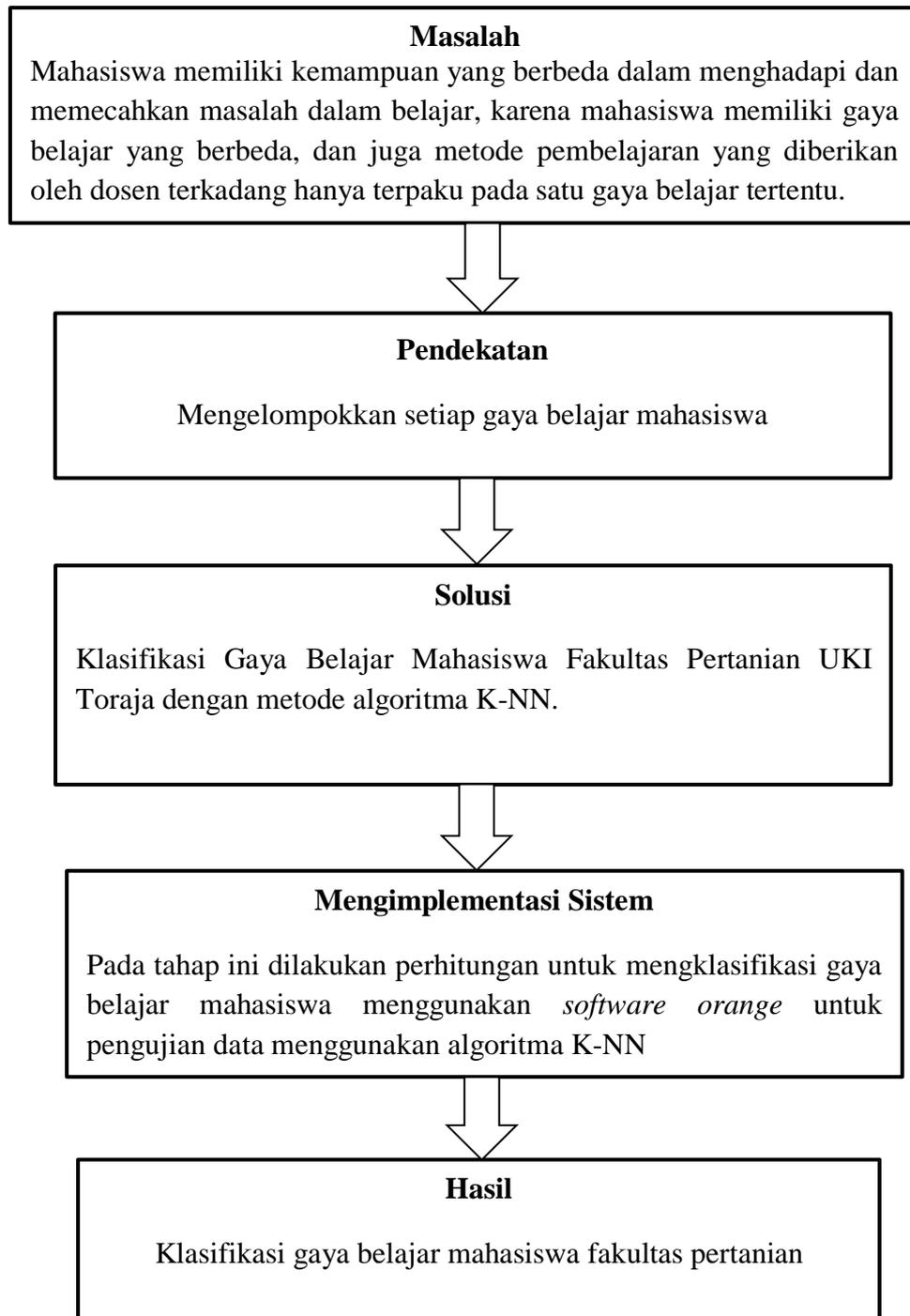
	Aktual : Positif (1)	Aktual : Negatif (0)
Prediksi : Positif (1)	TP	FP
Prediksi : Negatif (0)	FN	TN

Adapun rumus dalam menentukan akurasi dari tabel *Confusion matriks* adalah:  $Accuracy = (TP+TN) / (TP+FP+FN+TN)$

### 2.2.8 Software Orange

*Orange* adalah *software open source* untuk pengolahan data *analytics* atau data *mining*, dibandingkan dengan *software data mining* lainnya, *orange* lebih unggul dalam hal visualisasi atau yang biasa disebut *visual programming*. *Orange* menyediakan banyak *widget* yang diletakan pada canvas atau *drawing board* kemudian dapat dihubungkan dengan *widget* lainnya, dengan media canvas ini maka akan memudahkan pemakai bermain dengan data dan melakukan proses data analisis secara intuitif, setiap *widget* yang ada mempunyai fungsi masing-masing dan dapat menerima *input* dan mengeluarkan *output*.

### 2.3 Kerangka Pikir



Gambar 2.3 Kerangka pikir