**BAB II**

**KAJIAN PUSTAKA**

* 1. **Kajian Pustaka**

Seperti yang dijelakan pada sub-bab sebelumnya mulai dari karakteristik butir hingga *computerized adaptive testing* mulai dari 1970 an hingga kini, sudah ada beberapa pendekatan – pendekatan metode *adaptive* dari CAT. Mulai dari motode sederhana yang hanya menakin tingkat kesulitan jika benar dan menurunkan jika salah, hingga metode dengan memanfaatkan disiplin ilmu lain sepeti kecerdasan buatan, misalnya seperti penelitian ini menggunakan logika fuzzy.

Sejauh yang penulis cari, dari berbagai media. Penelitian sebelumnya tentang pengembangan *computerized adaptive testing* dengan metode fuzzy ditemukan pada penelitian sodara Haryanto pada 2009, pada karya disertasi beliau untuk gelar Doktor Pendidikan pada program studi Penelitian dan Evaluasi Pendidikan. Disamping penelitian tersebut, juga ditemukan penelitian yang sama yaitu dari sodara Diah Kusumawati, Andharini Dwi C dan Muhammad Fuad pada tahun 2014 akan tetapi setelah penulis kaji penelitian mereka meiliki kesamaan dengan penelitian bapak Haryanto, dan mengingat tahun penelitianya serta kelengkapan data dari penelitian tersebut, penulis hanya akan membandingkan dengan penelitian dari sodara Haryanto.

Penelitian saudara Haryanto memang menjadi rujukan utama bagi penulis untuk mengembangkan penelitian ini. Penulis telah melakukan kajian mendalam terhadap naskah disertasi beliau yang berjudul “Pengembangan Computerized Adaptive Testing (CAT) dengan Algoritma Logika Fuzzy”. Kemudian penulis juga telah melakukan kajian mendalam melalui berbagai literasi yang valid tentang evaluasi pendidikan khususnya tentang item response butir, teori klasik dan pengujian adaptif. Kemudian penulis merumusan beberapa hal yang menjadi pembeda antara penelitian penulis dan penelitian saudara Haryanto. Berikut didpaparkan beberapa perbedaanya dalam table 2.3

**Table 2.3 perbedaan penelitian**

|  |  |  |
| --- | --- | --- |
| **no** | **Penelitian Saudara Haryanto** | **Penelitian Agung Nur Hidayat** |
| 1. | Produk penelitina ini hanya menghasilkan perangkat lunak untuk menyajikan test secara adaptive | produk penelitian ini menghasilkan tool untuk kalibrasi soal yang outputnya adalah daya pembeda soal, tigkat kesulitan soal dan kluster soal. |
| 2. | Untuk membangun bank soal soal harus di ujikan dengan paper base test kemudian di kalibrasi secara manual baru kemudian di input ke dalam sistem beserta karakteristiknya | Selain bisa menyajikan tes secara adaptif perangkat lunak ini juga bisa menyajikan tes secara klasik, kemudian hasilnya bisa digunakan sebagai perhitungan untuk analisa karakteristik butir soal. |
| 3. | Output implementasi metode fuzzy di implikasikan lengsung ke tingkat kesulitan butir, sehingga kemungkinan untuk tidak ada ada nya butir soal yang presisi pas dengan output logika fuzzy cukup besar, terutma jika koleksi soal kurang hiterogen. | output logika fuzzy dimplikasikan ke cluster soal, yang bebararti hasil keputusan lebih fleksible menyesuaikan butir soal yang ada. Kunci dari pengklsuter adalah mengklusterkan soal yang “ada” menjadi sejumlah kluster yang diinginkan. Hal ini menyebabkan sistem lebih fleksible terhadap kondisi butir soal |
| 4. | Asumsi pada penelitian ini adalah kemampuan siswa tidak akan turun jika salah menjawab soal, tapi mencarikan soal lain yang memiliki tingkat kesulitan yang sama | Asumsi pada penelitian ini adalah kemampuan peserta bisa di asusmsikan turun saat dia menjawab pertanyaan salah. Sehingga metode adaptive pada penelitian ini akan menurunkan tingkat kesulitan soal jika peserta menjawab salah dengan cara membalik nilai keanggotaan himpunan fuzzy output pada *rule*-nya. |
| 5. | *Skroing* pada penelitian ini bersifat general artinya metode *populating skore* yang dikembangkan adalah umum. Bisa dikembangkan ke pengukuran kemampuan, seleksi maupun tujuan tes lainya | *Skoring* pada penelitian ini memfokuskan pada fungsi dan tujuan ujian sebagai seleksi untuk peserta tes. Sesuai objek implementasi dimana pengujian adaptive untuk menyeleksi peserta didik dari Sekolah Polisi Negara Selopamioro yang akan diberi beasiswa untuk melanjutkan pendidikan ke AKABRI. |
| 6. | Perangkat dikembangkan dengan bahasa VB. | Perangkat lunak dikembangkan dengan dengan 2 platform, yaitu web untuk guru dan admin lalu desktop untuk siswa. |
| 7. | Data base menggunakan *local server* | Berbasis internet / hosting server. |

* 1. **Karakteristik Butir**

Karakteristik butir merupakan parameter butir soal, yang menunjukan kualitas dari butir soal tersebut. Karakteristik butir digunakan untuk menentukan layak atau tidaknya sebuah soal digunkan. Karakteristik butir soal didapatkan dengan melakukan perhitungan dengan menggunakan rumus tertentu terhadap beberapa jawaban/respon dari peserta ujian yang mengerjakan butir soal tersebut. Dalam Teori Responsi Butir karakteristik butir terdiri dari tingkat kesulitan soal (*item difficulty*), daya pembeda soal (*item discrimination*) dan faktor kebetulan menjawab (*pseudo guesing*). Dalam penelitian ini seperti dijelaskan dalam batasan masalah karena yang berhubungan langsung dengan penyajian soal yang *adaptive* hanya 2 parameter logistik saja yaitu tingkat kesulitan soal dan daya pembeda soal. Berikut penjelasan lebih lanjutnya.

* + 1. **Tingkat kesulitan soal / *item difficulty***

Tingkat kesulitan soal atau *item difficulty* adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang bisanya dinyatakan dalam bentuk indeks. Range indeks tersebut biasanya berkisar antara 0,00 – 1,00. Soal dengan indeks kesukaran 0,00 berarti tidak ada peserta tes yang berhasil menjawab dengan benar, sebalik nya soal dengan indeks kesukaran 1,00 berarti semua peserta tes menjawab benar butir soal tersebut. Jika di tuliskan dalam rumus :

……. (2.1)

|  |
| --- |
| b = tingkat kesulitan soal |
| Np = jumlah siswa yang menjawab benar |
| N = jumlah jawaban siswa |

Klasifikasi tingkat kesulitasn soal menurut Nitko(1996)[5] seperti yang di tulis dalam bukunya dibagi menjadi 3, bisa di lihat pada table 2.1.

**Tabel 2.1 klasifikasi tingkat kesulitan soal**

|  |  |
| --- | --- |
| **Range Kesukaran / *b*** | **Klasifikasi soal** |
| 0,00 – 0,30 | Sukar |
| 0,31 – 0,70 | Sedang |
| 0,71 – 1,00 | Mudah |

* + 1. **Daya Beda Butir / *item discrimination***

Daya beda butir soal adalah nilai indeks yang menunjukan tingkat kemampuan butir soal untuk membedakan kelompok peserta ujian yang berkemampuan tinggi dengan dan berkemampuan rendah. Sebuah soal yang baik harus bisa membedakan mana siswa yang bisa dan mana siswa yang tidak bisa. Jadi butir soal yang memiliki daya beda yang baik akan terjawab benar oleh siswa yang memiliki kemampuan tinggi dan sebaliknya akan terjawab salah oleh siswa yang memiliki tingkat kemampuan rendah. Seperti tingkat kesulitan soal, daya beda juga di cari dengan menggunakan perhitungan dari analisa respon dari peserta ujian sebelumnya. Daya beda dapat di hitung dengan rumus sebagai berikut :

……….(2.2)

|  |
| --- |
| a = daya beda soal |
| BA= jumlah jawaban benar pada kelompok atas |
| BB = jumlah jawaban benar pada kelompok bawah |
| N = jumlah jawaban |

Untuk klasifikasi daya beda, pada penelitian ini penulis mengacu pada penelitian Crocker & Algina (1986)[6] yang dipaparkan pada tabel 2.2

**Tabel 2.2 klasifikasi daya beda soal**

|  |  |
| --- | --- |
| **Range Daya beda / *b*** | **konklusi** |
| b ≥ 0,4 | baik |
| 0,3 ≤ b < 0,4 | cukup baik |
| 0,2 ≤ b < 0,3 | cukup |
| b < 0,2 | buruk |

Pembagian kelompok baik atas maupun bawah diambil dari 25% - 33%[7] dari jumlah peserta yang mengerjakan butir soal tersebut yang memiliki nilai atau skor tertinggi untuk kelompok atas dan nilai atau skor terendah untuk kelompok bawah. Pada hakikatnya range daya beda adalah tidak negatif hal yaitu butir soal tersebut tidak dijawab benar justru lebih banyak oleh kelompok bawah. Hal ini berarti butir soal yang memiliki daya beda negatif justru tidak mencerminkan tingkat kemampuan siswa. Berbeda dengan tingkat kesulitan soal yang berbanding terbalik dengan tingkat kemampuan siswa, daya beda soal sebanding dengan tingkat kemampuan siswa, semakin besar daya beda soal maka semakin layak butir soal tersebut untuk disajikan.

* 1. **Teori Responsi Butir**

Pada pengujian teori klasik seperti *paper base test* dengan model urutan soal yang tetap akan menurunkan validitas dari ujian, misalnya kemungkinan *cheating* dari peserta tes. Selain itu karena keajegan tersebut maka nilai dari ujian tidak bisa menggambarkan kemampuan maksimal dari peserta ujian. Dalam ujian teori klasik / *classical test theory* dengan kondisi keajegan seperti diatas taraf sukar butir bergantung / *dependent* pada kemampuan responden. Pada peserta ujian yang memiliki tingkat kemampuan tinggi butir soal yang sama akan lebih mudah dibanding peserta dengan kemampuan yang lebih rendah. Jika untuk tujuan tes konvensional tentu tidak masalah, hanya saja dengan perkembangan teori dan ilmu dalam evaluasi pendidikan, sebuah ujian juga harus bisa mencerminkan kemampuan dari siswa dan bisa menggambarkan perkembangan kemampuan peserta tes dari waktu ke waktu.

Dengan perkembangnya asusmsi diatas dalam dunia evaluasi pendidikan, diawali tahun 1916 oleh Binet dan Smon. Kemudian pada tahun 1936, penelitian Ricahard menemukan keterkaitan antara karakteristik butir dengan perhitungan estimasi kemampuan peserta ujian. Baru kemudian pada 1952, ditemukan rumus model perhitungan peluang terjawab benar suatu butir soal berdasarkan analisis karakteristik butir soal yang kemudian disebut dengan “Parameter Logistik” yang terdiri dari tingkat kesulitan soal (*item difficulty*), daya beda (*item discrimination*) dan tebakan semu (*pseudo guessing*). Kemudian teori analisis tersebut yang kemudian dikenal dengan nama Teori Responsi Butir / *Item Response Theory* yang disingkat dengan IRT. Berdasarkan parameter logistic yang dipakai dalam menghitung peluang terjawab benar IRT memiliki 3 model yang sudah disepakati dan dibakukan [3] yaitu :

* 1 Parameter Logistik (L1P), dalam model ini hanya menggunakan tingkat kesulitan soal saja.
* 2 Parameter Logistik (L2P), dalam model ini menggunakan parameter logistik tingkat kesulitan soal dan daya pembeda soal.
* 3 Parameter Logistik (L3P), dalam model ini menggunakan semua parameter logistic yaitu tingkat kesulitan, daya pembeda soal serta tebakan semu atau tingkat kebetulan betul / *pseudo guessing.*

Seperti yang dijelaskan pada batasan masalah dalam penelitian ini penulis menggunakan model 2 parameter logistik. Sebab untuk tebakan semu menyangkut sektor psikometri untuk pengukuranya terutama untuk soal essay. Sehingga dikawatirkan bahasan dari penelitian ini terlalu meluas. Kemudian dalam IRT model L2P untuk pengukuran peluang terjawab benar subuah butir soal dirumuskan sebagai berikut :

……….(2.3)

|  |
| --- |
| = probabilitas siswa untuk menjawab dengan benar pada tingkat kemampuan **θ** |
| ***e*** = bilangan natural (2,718) |
| a = daya beda soal |
| b = tingkat kesulitan soal |
| **Θ =** tingkat kemampuan siswa |

Sesuai kaidah IRT sebuah soal yang baik harus memiliki peluang terjawab benar yang sama oleh setiap peserta tes dengan demikian sebuah hasil tes bisa menggambarkan kemampuan maksimal dari peserta tes tersebut. Sementara disisi lain kemampuan peserta tes berbeda – beda. Dari situlah konsep *adaptive test* kemudian mulai dikembangkan. Sekarang yang menarik untuk di kembangkan adalah bagaimana metode adaptif nya.

* 1. **Computerized Adaptive Testing**

Sesuai pemaparan diatas pada sub bab IRT, adalah yang mendasari konsep dari *adaptive testing* atau pengujian adaptif. Yaitu sebuah ujian yang butir soalnya(penyajian) mampu menyesuaikan dengan tingkat kemampuan peserta ujian. Disisi lain perkembangan yang cukup pesat dalam dunia komputer pada tahun 1970 an dengan di temukanya komputer generasi ke 3 dan penggunaan integrated circuit yang memungkinkan komputer untuk dapat melakukan proses manajemen data yang cukup kompleks. Lalu mulai tahun 1989[4] implementasi IRT dalam penyajian tes mulai pesat di kembangkan, dan model pengembangan tersebut yang kemudian dikenal dengan istilah *Compterized Adaptive Testing* yaitu penyajian ujian yang adaptif menyesuaikan kemampuan peserta.

Kemudian yang perlu dan menarik dikembangkan dalam dunia pendidikan terkait CAT adalah bagaimana metode ”*adaptive*” yang digunakan dari CAT itu sendiri. Dalam pengembangan CAT ada beberapa poin yang perlu diketahui yaitu :

* Metode pemilihan butir soal, seperti yang dijelaskan sebelumnya adaptif dalam CAT adalah menyesuaikan dengan tingkat kemampuan peserta ujian. Oleh karena itu metode pemilihan butir soal adalah salah satu hal pokok dalam pengembangan CAT.
* *Stoping rule*, adalah kondisi dimana CAT dihentikan. Ini sangat bergantung dengan tujuan tes dan metode pemilihan butir soal yang digunakan.

Pada penelitian ini motode pemilihan butir soal menggunakan kombinasi logika Fuzzy yang outputnya diimplikasikan rule baru untuk menghasilkan kluster soal.

* 1. ***Artificial Intellegence /*  Kecerdasan Buatan**

*Artificial Intellegence* (AI) dapat didefinisikan sebagai cabang dari ilmu komputer yang berkaitan dengan otomatisasi perilaku cerdas[8]. Ruang lingkup dari AI sendiri cukup luas di antaranya : (1) Pemrosesan Bahasa Alami, (2) Komputasi Neural, (3) Sistem Pakar, (4) Logika Samar/ Fuzzy, (5) Pemahaman Ucapan, (6) Komputer Visi, (7) Intellegent Tutoning, (8) Pemrograman Otomatis, (9) Robotika.

AI sendiri terdiri dari tiga unsur pokok[8] yaitu : (1) Knowledge Base (KB) dalam hal ini adalah pengetahuan tentang evaluasi pendidikan, (2) Data Base (DB) atau basis data yaitu sebuah sistem untuk menyimpan data olahan, dalam sistem ini menggunakan mysql database, (3) Inference Engine (IE) atau mesin inferensi, yang dalam sistem ini menggunakan Logika Fuzzy.

* 1. **Logika Fuzzy / *Fuzzy Logic***

Logika Fuzzy pertama kali ditemukan dan dipublikasikan pada 1965 oleh Lotfi A. Zadeh melalui papernya tentang *fuzzy set* yang berisi sebuah teori tentang ketidaktepatan (*imprecise*) dan ketidakpastian (*uncertainty*). Zadeh adalah seorang ahli matematika yang menempuh pendidikan di the University of Tehran, Tehran, Iran untuk gelar sarjana dan melanjutkan S2 nya di Massachusetts Institute of Technology (MIT) Cambridge, Massachusetts lalu gelar Ph.D. di Columbia University, New York[9].

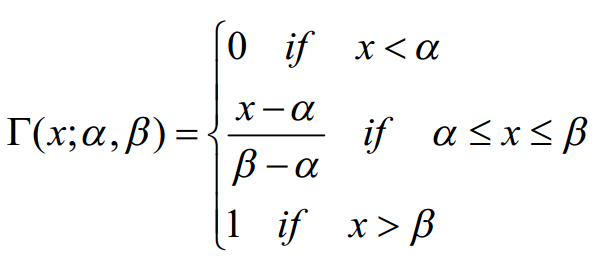
Logika fuzzy adalah algoritma matematika untuk menangani informasi yang tidak pasti, ambigu, dan perkiraan[10]. Logika fuzzy umumnya diterapkan pada masalah-masalah yang mengandung unsur ketidakpastian (*uncertainty*), ketidaktepatan (*imprecise*), *noisy*, dan sebagainya. Logika fuzzy adalah sebuah penalaran monoton untuk melakukan sebuah estimasi terhadap sebuah ketidakpastian. Ketidakpastian yang dimaksud dalam penelaran ini adalah adanya kemungkianan dari konklusi atau output kasus, dalam range value tertentu. Yang kemudian disebut dengan himpunan fuzzy baik input maupun output.

*Fuzzy logic* sendiri memiliki beberapa jenis model control, yaitu model atau tahapan dari pengumpulan himpunan fuzzy hingga melakukan output(*defuuzifikasi*). Model tersebut misalnya Model Kontrol Mamdani, Model Kontrol Sugeno dan Model Kontrol Tsukamoto[11]. Dalam penelitian ini model yang digunkan adalah *Tsukamoto* digunakan sebagai sistem inferensi untuk menentukan estimasi tingkat kemampuan siswa yang kemudian hasil defuzifikasi tersebut diimplikasikan untuk mengkasilkan tingkat cluster soal yang akan disajikan selanjutnya, yang memiliki beberapa tahapan yaitu :

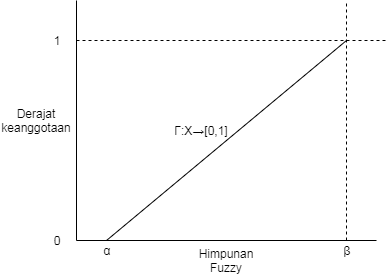
* + 1. **Pembentukan Himpunan Fuzzy / *Fuzzyfikasi***

seperti yang dijelaskan diatas logika fuzzy adalah estimasi terhadap ketidakpastian dan ketidaktepatan dari range yang sudah ditetapkan. Pada step ini adalah pembentukan variabel input dan variabel output. Pembentukan himpunan fuzzy didasarkan pada fungsi keanggotaan fuzzy. Berikut bebrapa macam fungsi keanggotaan yang digunakan dalam penelitian penelitian ini untuk membentuk keanggotaan fuzzy keanggotaan atau yang lebih dikenal dengan *fuzzy set*[11]

* *Γ-shaped fuzzy set* atau dalam istilah bahasa indonesia disebut fungsi linear naik, yaitu fungsi linier naik dengan satu parameter Γ:X→[0,1] dengan perhitungan sebagai berikut :

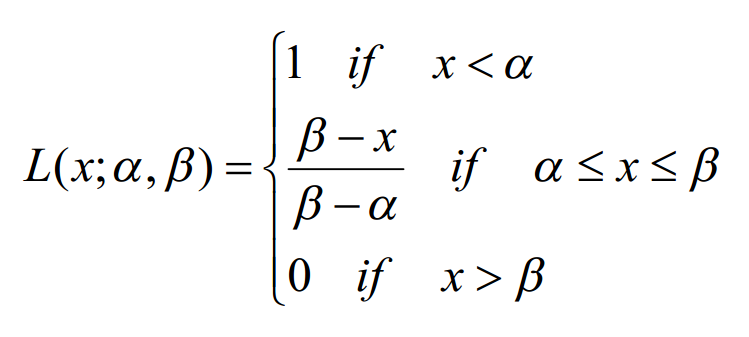
……….(2.4)

|  |  |
| --- | --- |
| a | : nilai parameter |
| β | : nilai maksimal pada himpunan fuzzy |
| α | : nilai minimal pada himpunan fuzzy |
| x | : nilai estimasi |

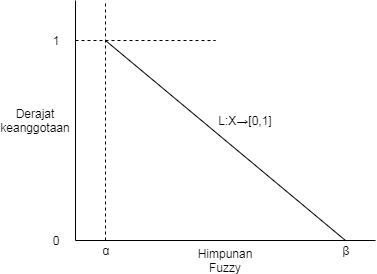


**Gambar 1 Graph Γ-shape fuzzy set**

* *L-shaped fuzzy set* atau dalam bahasa indonesia disebut fungsi linier turun yaitu fungsi linier penurunan naik dengan satu parameter L : :X→[0,1] dengan perhitungan sebagai berikut :

……….(2.5)

|  |  |
| --- | --- |
| a | : nilai parameter |
| β | : nilai maksimal pada himpunan fuzzy |
| α | : nilai minimal pada himpunan fuzzy |
| x | : nilai estimasi |



**Gambar 2 Graph L-shape fuzzy set**

* + 1. **Implementasi aturan / *rule*, yaitu penerapan perhitungan**

Pada tahap ini dilakukan implementasi funsi basis aturan berdasarkan *knowledge base,* yang sudah dirumuskan sebelumnya. Pola aturan ini biasanya berupa implikasi yang didalamnya operasi – operasi logika fuzzi seperti *and , or* dan lainya, yang akan digunakan untuk mengoperasikan variabel input. Pada penelitina ini semua rule implikasi menggunkan operasi logika *or.* Pada model tsukamoto operasi *or* menggunakan fungsi implikasi maksimal. Implementasi aturan digunakan sebagai dasar untuk melakukan inferensi*,* yaitu tahap selanjutnya.

* + 1. ***Inferension /* Inferensi**

Penegasan keputusan berdasarkan komposisi aturan. Komposisi aturan adalah semua aturan / *rule* yang sudah ditetapkan berdasar *knowledge base.* Pada tahap ini dihasilkan bobot dari tiap aturan yang kemudian akan dilakukan defuzzyfikasi*.*

* + 1. ***Defuzzyfikasion /* Defuzifikasi**

input dari defuzifikasi merupakan bobot yang dihasilkan dari tiap – tiap rule pada tahap inferensi. Kemidian pada tahap ini dilakukan rerata bobot. Output dari tahap ini adalah yang menjadi output dari logika fuzzy yaitu nilai hasil estimasi yang menjadi kodomain dari domain himpunan variabel fuzzy output.

* 1. **Data Mining**

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan atau *machine learning* untuk mengekstraksi dan mengindentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database[13]. Pada perinsipnya data mining adalah tehnik olah data tingkat lanjut.

Data dapat dilihat sebagai bentuk dari evolusi alami dari teknologi informasi. Dunia industri sistem data base mengalami perubahan yang cukup pesat terutama dari segi fungsionaliasnya. berawal hanya sebagai tempat penyimpanan primitif pada 1960 dan sebelumnya), kemudian mulai mengenal manajemen database pada 1970 hingga 1980 mulai ditemukan teknik pengolahan data lanjut, termasuk salah satunya data mining[14].

Teknik Data mining sendiri ada berbagai jenis berdasarkan fungsi dan mining patternnya. Dalam sub bab ini penulis hanyaakan menyebutkan sekaligus sedikit menjelaskan tentang klasifikasi teknik data mining[14], sedikitnya bisa dikalsifikasikan sebagai berikut :

* *Assosiation Pattern,* pola ini mendiskripsikan teknik pengolahan data kaidah assosiasi diantara kumpulan data. Penggunaan sering digunakan untuk mengelopokan barang sesuai dengan pola pembelian konsumen. Misalnya algoritma A priory.
* *Cluster Pattern,* pola ini membagi kumpulan data kedalam grup – grup dengan jumlah tertentu berdasarkan tingkat kemiripan yang menjadi parameter. Kaidahnya adalah memabagi antar grup dengan perbedaan sebesar mungkin dan mengelompokan anggota grup semirip mungkin. Algoritma terapanya misal *fuzzy c-Means*  dan yang digunakan pada penelitian ini untuk mengklusterkan soal berdasarkan kesulitanya adalah algoritma *k-Means.*
* *Classification pattern,* pola ini menjadi fungsi klasifikasi juga biasa disebut dengan *classifier.* Pengklasifikasian data memiliki 2 step yaitu *learning* dan *classification.*
* *Sequence pattern,* pola ini mengkombinasikan analisa assosiasi dan wakti diantara kumpulan data. Untuk menemukan pola ini, kita tidak hanya membutuhkan pengetahuan tentang kejadian yang terjadi tapi juga tentang waktu yang akurat dari kejadian tersebut terjadi.
  1. ***K-Means Clustering***

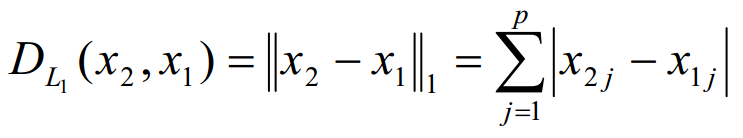
Seperti dijelaskan diatas *cluster analysis* adalah metode pembelajaran formal dan algoritma untuk mengelompokan atau mengklusterkan objek berdasarkan pengukuran atau karakteristik atau kesamaanya[16]. Perkembangan teknik *clustering* sendiri cukup pesat dan cukup luas. Ahli taksonomi, sosial, psykologi, statistik, matematik, *engineers,* ahli komputer, kesehatan dan bidang lainya yang mengkoleksi data dan pemrosesannya memiliki andil dalam perkempangan metode pengklusteran data. *Data clustering* pertama kali di kenalkan pada 1954 pada artikel tentang data antropologi. *Data clustering* juga dikenal dengan nama *Q-analysis, typology, clumping and taxonomy* tergantung dimana metode itu digunakan[16].

Salah satu algoritma dalam *cluster analysis* dan yang digunakan dalam penelitian ini adalah *k-Means Clustering*. *K-Means* merupakan algoritma pengklusteran sederhana pertama yang di publikasikan. Algoritma ini pertama dienalkan pada 1955[16]. Hingga kini perkembangan metode pengklusteran sendiri sudah sangat luas, kendati demikian *k-Means* sendiri masih cukup populer digunakan di seluruh dunia dan di beberapa disiplin ilmu khususnya untuk pengklusteran data. Hal yang perlu ditekankan atau menjadi kaidah pada algoritma ini adalah pada kemiripan antar data dari data yang tersedia.

Untuk tahapan pengklusteran data pada algoritma *k-Means* adalah dengan mengelompokan data pada sebuah titik central yang disebut dengan *centroid. Centroid* merupakan sebuah nilai diantara domain data yang diklusterkan. *Centroid* akan menjadi *mean* atau rerata dari suatu kluster. Posisi data terhadap kluster akan di tentukan oleh kedekatanya dengan centroid terdekat. Jumlah *centroid* menetukan jumlah kluster yang akan dihasilkan oleh proses pengklusteran. Kaidahnya jumlah kluster harus lebih sedikit dari jumlah data unik dari kumpulan data yang akan diklusterkan. Karena tiap kluster pasti akan memiliki member data.

Untuk tahapan *k-Means* sendiri ada beberapa tahap hingga terbentuk kluster berdasarkan kedekatan dengan centroid sebagai berikut[17] :

1. Tentukan jumlah kluster, dalam penelitian ini ditentukan menjadi 4 kluster yang akan menjadi tingkat kesulitan dari soal.
2. Alokasikan data kedalam kluster secara random, pada penelitian ini inisialisasi pengaloasian kluster data awal adalah berdasarkan kedekatan terhadap 4 data unik terendah dalam kumpulan data tersebut.
3. Hitung centroid / rata – rata dari data yang ada di masing –masing kluster.
4. Alokasikan tiap – tiap data kedalam centroid terdekat, untuk pengukuran jarak yang dipakai pada penelitian ini adalah *Manhattan/City Block* yaitu dengan rumus :

……….(2.6)

|  |  |
| --- | --- |
| DL | : jarak data |
| *x* | : data |
| *p* | : dimensi data |
| |.| | : nilai absolut |

1. Lakukan iterasi dari langkah 2 apabila belum memenuhi *stoping rule* dibawah ini :

* Iterasi sudah melebihi 2500 kali. Hal ini dikarenakan keterbatasan hardware server saat menerima requet time php.
* Tidak ada data yang berpindah kluster.
  1. **Analisis Kelemahan Sistem Lama**

Metode yang digunakan dalam analisis kelemahan sistem lama adalah metode PIECES (*Perfomance, Information, Economic, Control, Efficiency, Service*).

Untuk mengidentifikasi masalah, harus dilakukan analisis terhadap kinerja, informasi, ekonomi, keamanan aplikasi, efisiensi dan pelayanan pelanggan. Paduan ini dikenal dengan analisis PIECES. Dari analisis ini biasanya didapat beberapa masalah utama. Hal ini penting karena biasanya yang muncul di permukaan bukan masalah utama, tetapi gejala dari masalah utama saja.

* + 1. **Analisis Kinerja (Perfomance)**

Analisis kinerja ditunjukan untuk mengetahui tingkat kerja dari sebuah sistem apakah kinerja dari sistem tersebut telah berjalan sesuai dengan yang diharapkan atau sudah mencapai sasaran yang diinginkan.

* + 1. **Analisis Informasi (Information)**

Sebuah sistem informasi yang baik dan akan menghasilkan informasi akurat, relevan dan tepat waktu. Akurat berarti informasi yang dihasilkan terbebas dari kesalahan dan tidak menyesatkan, sedangkan relevan berarti informasi tersebut memiliki nilai bagi penggunanya, dan tepat waktu berarti informasi harus ada ketika dibutuhkan. Ketiga kriteria tersebut merupakan syarat dari informasi yang baik bagi sebuah perusahaan atau organisasi yang akan dijadikan sebagai dasar dari penggambilan keputusan.

* + 1. **Analisis Ekonomi *(Economy)***

Motif ekonomi mungkin merupakan salah satu pertimbangan dari alasan mengapa diperlukan pengembangan sistem. Harapan sebuah perusahaan atau organisasi terhadap sistem yang baru adalah dukungan terhadap proses manajerial perusahaan yang lebih efesien. Sehingga adanya pemborosan waktu dan alat-alat yang dapat mengakibatkan pembengkakan biaya pada sistem sebelumnya (sistem lama) dapat dikurangi semaksimal mungkin.

* + 1. **Analisis Keamanan *(Control)***

Aktivitas sebuah perusahaan atau organisasi perlu mendapat perhatian dan control yang terus menerus agar tidak terjadi penurunan kinerja dibawah standar yang telah ditetapkan. Hal ini untuk mengurangi dan mencegah atau mendeteksi kesalahan sistem, menjaga keamanan data dan kecurangan yang akan terjadi. Pengendalian sebuah sistem sangat diperlukan untuk mengurangi dan mencegah hal-hal yang merugikan perusahaan dan organisasi itu sendiri. Dengan adanya control maka tugas atau kinerja yang mengalami kendala dapat diperbaiki.

* + 1. **Analisis Efisiensi *(Efficiency)***

Effisiensi pada sebuah sistem informasi menyangkut hal bagaimana menghasilkan output atau informasi sebaik mungkin dengan input yang diberikan, sehingga informasi yang dihasilkan sesuai dengan kebutuhan bagi pihak-pihak yang memerlukannya. Selain itu efisiensi juga berkaitan dengan bagaimana sebuah sistem tidak melakukan pemprosesan secara berlebihan, dan usaha yang dikeluarkan untuk melakukan tugas-tugas juga tidak berlebihan.

* + 1. **Analisis Layanan *(Service)***

Untuk menilai kualitas dari sebuah sistem adalah dengan melihat segi pelayanannya. Penigkatan pelayanan terhadap sistem yang dikembangkan akan memberikan kehandalan terhadap konsistensi dalam pengolahan input dan output serta kehandalan dalam menangani pengecualian.

* 1. **Analisis Kebutuhan Sistem**

Tujuan dari fase analisis ini adalah untuk memahami dengan sebenar-benarnya kebutuhan dari sistem baru dan mengembangkan sebuah sistem yang mewadahi kebutuhan tersebut, atau memutuskan bahwa sebenarnya pengembangan sistem baru tidak dibutuhkan.

Untuk mempermudah analisis sistem dalam menentukan keseluruhan kebutuhan secara lengkap, maka analis membagi sistem kedalam dua jenis yaitu:

* + 1. **Kebutuhan Fungsional (Functional Requirement)**

Kebutuhan Fungsional adalah jenis kebutuhan yang berisi proses-proses apa saja yang nantinya dilakukan oleh sistem. Kebutuhan fungsional juga berisi informasi-informasi apa saja yang harus ada dan dihasilkan oleh sistem.

* + 1. **Kebutuhan Nonfungsional (Nonfunctional Requirement)**

Kebutuhan nonfungsional adalah tipe kebutuhan yang berisi property perilaku yang dimiliki oleh sistem seperti perangkat keras dan perangkat lunak apasaja yang digunakan dan seperti apa kinerja, keamanan serta informasinya berjalan.

* 1. **Analisis Kelayakan Sistem**

Ketika sistem selesai menyusun dokumen kebutuhan sistem, maka tahap desain bisa dimulai. Namun tidak semua kebutuhan sistem yang didefinisikan pada tahanapan analisis kebutuhan sistem layak untuk dikembangkan pada sistem informasi.

Untuk memastikan usulan sistem tersebut bisa diteruskan menjadi proyek yang menguntungkan maka proposal proyek harus dievaluasi kelayakannya dari segi :

* + 1. **Kelayakan Teknis**

Kelayakan teknis menyoroti kebutuhan sistem yang telah disususn dari aspek teknologi yang akan digunakan. Jika teknologi yang dikehendaki untuk pengembangan sistem merupakan teknologi yang mudah didapat, murah, dan tingkat pemakaiannya mudah, maka secara teknis usulan kebutuhan sistem bisa dinyatakan layak.

* + 1. **Kelayakan Operasional**

Untuk disebut layak secara operasional, usulan kebutuhan sistem harus benar-benar bisa menyelesaikan masalah yang ada di sisi pemesanan sistem informasi. Disamping itu, informasi yang dihasilkan oleh sistem harus merupakan informasi yang benar-benar dibutuhkan oleh pengguna tepat pada saat pengguna menginginkannya.

* + 1. **Kelayakan Ekonomi**

Kelayakan ekonomi berhubungan dengan *return on investment* atau berapa lama biaya investasi dapat kembali. Analisis kelayakan ekonomi juga akan mempertimbangkan apakah bermanfaat melakukan investasi ke proyek ini atau kita harus melakukan sesuatu yang lain.

* + 1. **Kelayakan Hukum**

Analisis sistem yang akan dikembangkan berdasarkan legalitas hokum yang digunakan pada stake holder sistem digunakan.

* 1. **Tahap Perencanaan / Desain**

Pada tahap desain, semua fungsi-fungsi dari suatu sistem dideskripsikan secara detail. Yang didalam melalui proses beberapa uji coba, sesuai pendekatan penelitian ini adalah *research and development.* Dimana umpan balik dari uji coba adalah input dari tahap perncanaan. Ada beberapa aktifitas yang perlu dilakukan pada tahapan ini, yaitu:

* + 1. **Membuat Pemodelan Sistem**

Untuk membuat pemodelan sistem tergantung dari jenis pemrograman yang digunakan. Apabila analis menggunakan pemrograman terstruktur maka dapat menggunakan *Flowchart* dan *DFD*. Sedangkan apabila analisis menggunakan pemrograman berbasis objek *(OOP)* maka dapat menggunakan *UML*.

* + 1. **Membuat Pemodelan Data**

Untuk membuat pemodelan data dapat menggunakan *Entity Relationship Diagram (ERD)* atau normalisasi.

* + 1. **Membuat *User Interface* Sistem**

Rancangan *user interface* ini digunakan sebagai acuan dalam proses implementasi agar sistem yang dibuat memiliki tampilan yang baik dan sistem dapat berjalan dengan baik.

* 1. ***Flowchart* Sistem**

Menurut Yakub (2012) flowchart adalah bagan yang menggambarkan urutan instruksi proses dan hubungan satu proses dengan proses lainnya menggunakan simbol-simbol tertentu. Flowchart digunakan sebgai alat bantu komunikasi dan dokumentasi. [18]

Flowchart Sistem merupakan bagan yang menunjukan arus pekerjaan secara keseluruhan dari sistem. Bagan ini menjelaskan urutan dari prosedur-prosedur yang ada didalam sistem.

Ada beberapa simbol yang digunakan saat membuat *flowchart*, yaitu :

|  |  |  |
| --- | --- | --- |
|  | : | Dokumen input atau output baik untuk proses manual ataupun mekanik. |
|  | : | Proses |
|  | : | Manual *input* / *keyboard* |
|  | : | Penyimpanan Data *(storage)* |
|  | : | Arah proses |

**Gambar 2. 3 Simbol Flowchart**

* + 1. **Entity Relationship Diagram (ERD)**

Menurut Al Fatta, Hanif (2007) *ERD* adalah gambar atau diagram yang menunjukan informasi dibuat, disimpan dan digunakan dalam sistem bisnis. *ERD* digunakan untuk menunjukkan aturan-aturan bisnis yang ada pada sistem informasi yang akan dibangun.[19]

Dalam *ERD* ada dua elemen penyusun, yaitu *entity* dan *relatonship*. *Entity* adalah sesuatu apa saja yang ada didalam sistem, nyata maupun abstrak dimana data tersimpan atau dimana data terdapat. Relationship adalah hubungan alamiah yang terjadi diantara entitas.[20]. Ada beberapa simbol yang digunakan saat membuat ERD, yaitu :

|  |  |  |
| --- | --- | --- |
|  | : | Entitas |
|  | : | Atribut |
|  | : | Relasi |
|  | : | Derajat kardinalitas 1 |
|  | : | Derajat kardinalitas N |

**Gambar 2. 4 Simbol ERD**

* + 1. **Data Flow Diagram (DFD)**

Menurut Ladjamudin (2005) diagram alir data merupakan model dari sistem untuk menggambarkan pembagian sistem ke modul yang lebih kecil. Salah satu keuntungan menggunakan diagram alir data adalah memudahkan pemakai atau user yang kurang menguasai bidang komputer untuk mengerti sistem yang akan dikerjakan[20]. Ada beberapa simbol yang digunakan saat membuat DFD, yaitu:

|  |  |  |
| --- | --- | --- |
| **Gane / Sarson** | **Yourdon / De Marco** | **Keterangan** |
|  |  | Entitas Eksternal |
|  |  | Proses |
|  |  | Penyimpanan Data |
|  |  | Aliran Data |

**Gambar 2. 5 Simbol DFD**

* 1. **Tahap Implementasi**

Tahap penerapan merupakan kegiatan memperoleh dan mengintegrasikan sumber daya fisik dan konseptual yang dibutuhkan oleh sistem agar dapat berjalan. Sesuai pendekatan penilitian yaitu *research and development* dan dengan model *prototype* untuk pengembangan perangkat lunaknya, maka dalam tahap ini ada beberapa hal yang dilakukan :

1. Membuat database sesuai dengan skema rancangan pada tahap desain.
2. Membuat aplikasi berdasarkan desain sistem.
3. Melakukan uji coba
4. Melakukan perbaikan sistem sesuai *feed back* hasil uji coba.
   1. **Definisi Basis Data**

Dikutip dari Utami, Ema dan Dwi Hartanto, Anggit (2012) basis data adalah kumpulan data yang saling berhubungan yang disimpan / diorganisasi secara bersama, dalam bentuk sedemikian rupa dan tanpa redudansi (pengulangan) yang tidak perlu supaya dapat dimanfaatkan kembali dengan cepat dan mudah untuk memenuhi berbagai kebutuhan. [21]

Menurut Kusrini (2007) basis data adalah kumpulan data yang saling berelasi. Basis data dapat didefinisikan dalam berbagai sudut pandang seperti berikut [22] :

1. Himpunan kelompok data yang saling berhubungan yang diorganisasi sedemikian rupa sehingga kelak dapat dimanfaatkan dengan cepat dan mudah.
2. Kumpulan data yang saling berhubungan yang disimpan secara bersamaan tanpa pengulangan *(redundancy)* yang tidak perlu, untuk memenuhi kebutuhan.
3. Kumpulan file / table / arsip yang saling berhubungan yang disimpan dalam media penyimpanan elektronik.
   * 1. **Komponen-komponen Basis Data**

Menurut Kusrini (2007) sistem basis data merupakan perpaduan antara basis data dan sistem manajemen basis data (DBMS). Komponen-komponen sistem basis data meliputi: Perangkat Keras, Sistem Operasi atau perangkat lunak untuk mengelola basis data, Database Management System (DBMS), pemakai, Aplikasi lain. [22]

1. Perangkat keras

2. Sistem operasi atau perangkat lunak untuk mengelola basis data

3. Database management system (DBMS)

4. Pemakai

5. Aplikasi lain