

BAB II

LANDASAN TEORI

2.1 Universitas Muhammadiyah Pringsewu Lampung

Universitas Muhammadiyah Pringsewu (UMPRI) Lampung merupakan lembaga pendidikan tinggi yang berdiri berkat proses penggabungan tiga Perguruan Tinggi Muhammadiyah di Kabupaten Pringsewu yaitu STKIP Muhammadiyah Pringsewu Lampung, STIE Muhammadiyah Pringsewu Lampung, dan STIKES Muhammadiyah Pringsewu Lampung. Penggabungan tiga perguruan tinggi Muhammadiyah yaitu STKIP Muhammadiyah Pringsewu Lampung, STIE Muhammadiyah Pringsewu Lampung, dan STIKes Muhammadiyah Pringsewu Lampung menjadi Universitas Muhammadiyah Pringsewu merupakan salah satu bentuk amanah yang dinyatakan pada acara Musyawarah Wilayah Muhammadiyah Tahun 2010.

2.2 Algoritma K-Means Clustering

Menurut (Wahyuningtyas, Putri, & Sutrisno,2018), menjelaskan bahwa : “K-Means Clustering adalah metode data cluster non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster dengan data yang hampir sama. Data yang memiliki karakteristik yang sama akan dikumpulkan dalam satu cluster, sedangkan data yang memiliki karakteristik berbeda akan dikumpulkan kedalam cluster yang lain. Karakteristik data yang terbesar dalam satu cluster memiliki tingkat variasi yang kecil”.

Metode *k-means* berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam suatu kelompok mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dasar algoritma *k-means* adalah sebagai berikut:

1. Tentukan nilai k sebagai jumlah klaster yang ingin dibentuk.
2. Inisialisasi k sebagai *centroid* yang dapat dibangkitkan secara *random*.
3. Hitung jarak setiap data ke masing-masing *centroid* menggunakan persamaan *Euclidean Distance* yaitu sebagai berikut:

$$d(P, Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2}$$

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroidnya*.

5. Tentukan posisi *centroid* baru (k).
6. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

2.3 Pengertian Data Mining

Menurut Alfa Saleh, (2015 : 208), menjelaskan bahwa : “Data Mining merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi – informasi yang nantinya dapat digunakan”.

2.3.1 Tahapan Data Mining

Data mining adalah salah satu proses yang ada didalam KDD menurut Fayyad dalam buku (Kusrini., 2009). Berikut merupakan rangkaian dari beberapa proses :

1. Data Selection (seleksi data)

Pemilihan atau seleksi data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *Knowledge discovery in databases* (KDD) dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional. Tujuan dari fase ini adalah ekstraksi dari kumpulan data yang besar menjadi data yang relevan dengan analisis data mining.

2. Pre-Processing / Cleaning (Pemilihan Data)

Proses *Preprocessing* mencakup antara lain membuang duplikasi data, memeriksa data yang *inkonsisten* dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*). Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi *eksternal*.

3. Transformasi (*Transformation*)

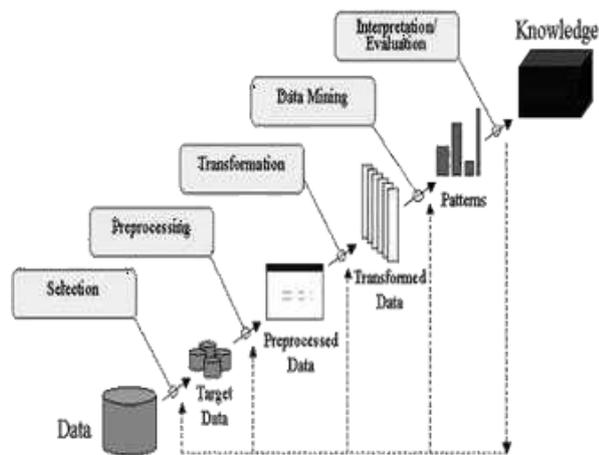
Pada fase ini yang dilakukan adalah mentransformasi bentuk data yang belum memiliki entitas yang jelas ke dalam bentuk data yang valid atau siap untuk dilakukan proses Data Mining.

4. Data Mining

Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode pencarian pengetahuan. Ini adalah langkah penting di mana teknik kecerdasan diterapkan untuk mengekstrak pola informasi yang berpotensi berguna dari data yang dipilih.

5. Interpretasi/Evaluasi (*Interpretation/Evaluation*)

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya. Pada fase terakhir ini yang dilakukan adalah proses pembentukan keluaran yang bersumber pada proses Data Mining Pola informasi (Nofriansyah., 2014) pada gambar 2.1 dibawah ini.



GAMBAR 2.1 Proses *data mining* (Nasari, 2015).

2.4 Clustering

Menurut Eko Prasetyo (2012:6), pengelompokkan data-data ke dalam sejumlah kelompok (cluster) berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada. Pengelompokkan data dibedakan menurut fakultas dan jurusan. Menurut struktur, pengelompokkan dibagi dua, yaitu hierarki dan partitioning. Dalam hierarki, satu data tunggal bisa dianggap sebuah kelompok, dua atau lebih. Pengelompokkan partitioning membagi setiap data hanya menjadi anggota satu kelompok. Menurut keanggotaan data dalam kelompok, dibagi menjadi dua, yaitu eksklusif dan tumpang-tindih. Dalam kategori eksklusif, sebuah data bisa dipastikan hanya menjadi anggota satu kelompok dan tidak menjadi anggota kelompok yang lain. Sedangkan kategori tumpang-tindih adalah metode pengelompokkan yang membolehkan sebuah data menjadi anggota lebih dari satu kelompok.

2.5 Perangkat Lunak Yang Digunakan

1. PHP

Menurut Dokumen Resmi PHP, PHP merupakan singkatan dari PHP Hypertext Preprocessor. PHP merupakan bahasa berbentuk skrip server side yang disisipkan pada

dokumen HTML dan ditempatkan serta di prosedur dalam server. Hasil yang didapatkan dikirimkan ke klien, tempat pemakai menggunakan browser. Secara khusus, PHP dirancang untuk membentuk web dinamis. Artinya, ia dapat membentuk suatu tampilan berdasarkan permintaan terkini. Misalnya, Anda bisa menampilkan isi database ke halaman web.

2. MYSQL

MySQL adalah Software Database Manajemen Sistem (DBMS), MySQL merupakan turunan salah satu konsep utama dalam database sejak lama, yaitu Structured Query Language (SQL). SQL adalah sebuah konsep pengoperasian database. MySQL merupakan software database yang paling populer digunakan untuk membangun aplikasi web yang menggunakan database sebagai sumber dan pengelola datanya, maka dari itu tidak heran bahwa kini MySQL sangat populer dikalangan pemrogram web yang menggunakan script PHP dan Perl baik dilingkungan platform opensource operating system, maupun windows.

3. Rapid Miner Studio

Rapid Miner adalah sebuah software untuk pengolahan data mining. Rapid Miner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik (Setiawan, 2017).

4. Microsoft Excel

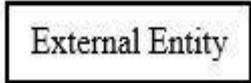
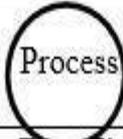
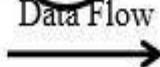
Dalam penelitian (Dhewy, 2018) Microsoft Excel adalah salah satu program dari Microsoft Word untuk mengaplikasikan lembar kerja (*worksheet*) atau lebih dikenal *spreadsheet program*. Di dalam Microsoft Excel terdapat formula yang masing-masing memiliki fungsi dalam beberapa bidang untuk menghitung nilai dan angka dengan cepat. Selain itu, Microsoft excel juga dapat menampilkan grafik dan gambar.

2.6 Perancangan Sistem

2.6.1 Diagram Konteks

Diagram konteks dibuat untuk menggambarkan sumber serta tujuan data yang akan diprosedur. Diagram konteks menjelaskan secara umum mengenai suatu prosedur yang menggambarkan ruang lingkup, rancang bangun sistem clustering data mahasiswa Berikut rancangan diagram konteks untuk sistem yang diusukan.

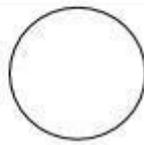
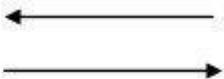
Tabel 2.1 Simbol Diagram Konteks

Simbol Diagram Konteks	Keterangan
	Pihak-pihak yang berada di luar sistem, tetapi secara langsung berhubungan dengan sistem dalam hal memberi data atau menerima informasi
	Didalam diagram konteks, berisi mengenai sistem yang akan dibuat
	Berisi data atau informasi yang mengalir dari satu pihak ke sistem dan sebaliknya

2.6.2 Data Flow Diagram

Data flow Diagram (DFD) digunakan untuk menggambarkan sistem yang di usulkan secara terstruktur dan jelas sesuai dengan analisis sistem menggambarkan mengenai pengolahan data yang ada di dalam sistem, seperti pengolahan kmeans clustering serta input data mahasiswa melalui admin.

Tabel 2.2 Simbol-simbol Data Flow Diagram

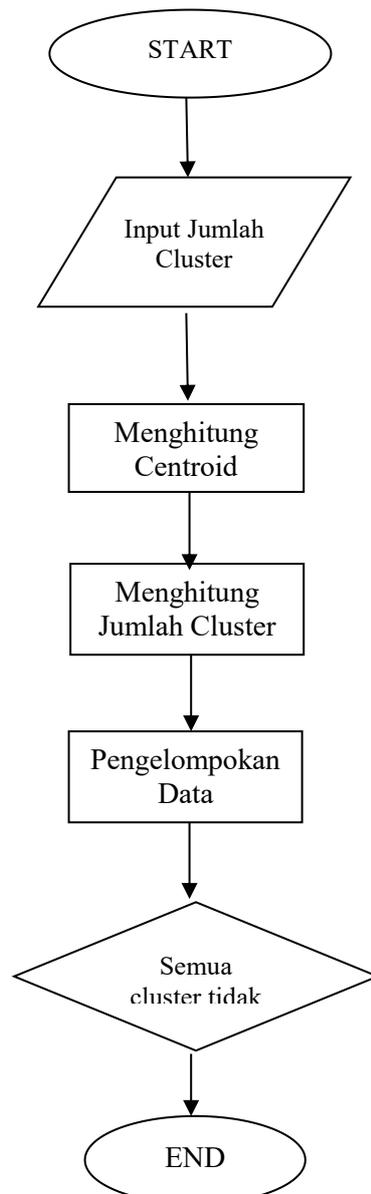
Simbol	Keterangan
	<i>External Entity</i> Simbol ini digunakan untuk menggambarkan asal atau tujuan data
	Proses Simbol ini menunjukan proses pengolahan atau tranformasi data
	<i>Data Flow</i> Simbol ini digunakan untuk menggambarkan aliran data yang berjalan
	<i>Data Store</i> Simbol ini menggambarkan <i>data Flow</i> yang sudah disimpan atau diarsipkan

2.6.3 Entity Relational Diagram (ERD)

Entity relationship diagram (ERD) adalah suatu model jaringan yang menggunakan susunan data yang disimpan dalam sistem secara abstrak, serta menguntungkan bagi professional sistem.

2.7 Flowchart

Flowchart adalah bagan-bagan yang mempunyai arus yang menggambarkan langkah-langkah penyelesaian suatu masalah. Flowchart merupakan cara penyajian dari suatu algoritma untuk clustering data mahasiswa. Berikut gambar flowchart pada gambar 2.2 dibawah ini



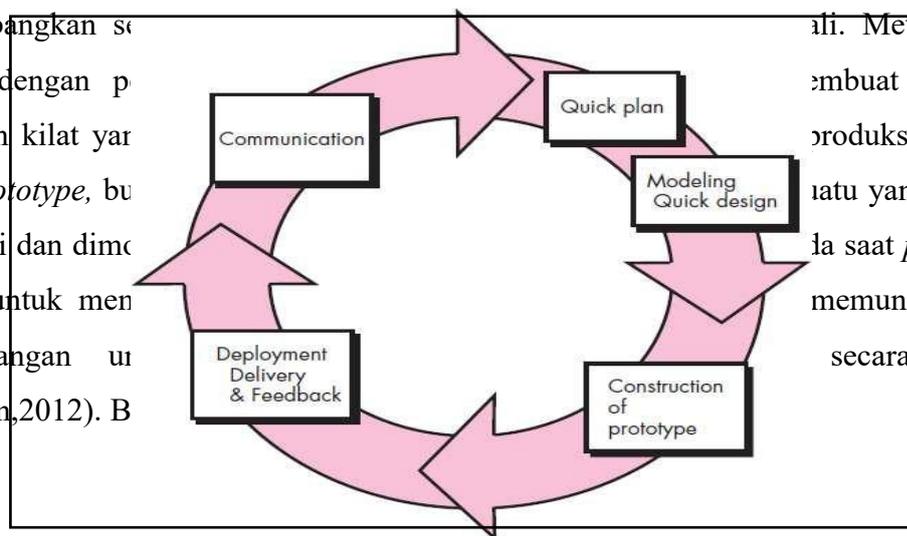
Yes

Gambar 2.2 Flowchart Clustering Data Mahasiswa

2.8 Metode Perangkat Lunak Yang Digunakan

2.8.1 Prototype

Prototype adalah proses pembuatan model sederhana *software* yang memungkinkan pengguna memiliki gambaran dasar tentang program serta melakukan pengujian awal. *Prototype* memberikan fasilitas bagi pengembang dan pengguna untuk saling berinteraksi selama proses pembuatan, sehingga pengembang dapat dengan mudah memodelkan perangkat lunak yang akan di buat. Metode ini cocok digunakan untuk mengembangkan software. Metode ini dimulai dengan pembuatan rancangan kilat yang benar *Prototype*, kemudian dievaluasi dan dimodifikasi di buat untuk meningkatkan pengembangan user (Pressman,2012). B



Gambar 2.3 Diagram Prototype

Tahapan-tahapan dalam metode *Prototype* :

1. Komunikasi (*Communication*) :pengumpulan data awal, yaitu komunikasi dengan klien dan *user* untuk menentukan kebutuhan.
2. Perencanaan Cepat (*Quick Plan*) : pembuaan perencanaan analisis terhadap kebutuhan pengguna.
3. Pemodelan Perancangan Cepat (*Modeling Quick Design*) : membuat rancangan desain program.
4. Pembentukan *Prototype* (*Construction of prototype*) : pembuatan aplikasi berdasarkan dari pemodelan desain yang telah dibuat.
5. Penyerahan Sistem dan Umpan Balik (*Development Delevery and Feedback*) : memproduksi perangkat secara benar sehinga dapat digunakan oleh pengguna.

2.9 Penelitian Terkait

Dalam penyusunan skripsi ini, penulis terinspirasi dan mereferensi dari penelitian-penelitian sebelumnya yang berkaitan dengan skripsi ini. Daftar penelitian ditunjukkan pada table 2.3:

Tabel 2.3 Penelitian Terkait

NO	NAMA	JUDUL	TAHUN	KETERANGAN
1.	ASRONI (1), HIDAYATUL FITRI (2), EKO PRASETYO (3)	Data Calon Mahasiswa Baru di Universitas Muhammadiyah Pringsewu Yogyakarta (Studi Kasus : Fakultas Kedokteran dan Ilmu Kesehatan, dan Ilmu Sosial dan Ilmu Politik)	2018	Sistem ini akan dibangun dan ditujukan untuk pengelompokan mahasiswa baru dan juga tentang perbandingan sebelum dan sesudah adanya pandemic.
2.	Miftachur Robani(1),	Algoritma K- Means	2018	Sistem ini bertujuan

	Achmad Widodo(2)	Clustering Untuk Pengelompokan Ayat Al Quran Pada Terjemahan Bahasa Indonesia		menghasilkan suatu sistem informasi yang dapat melakukan pengelompokan ayat Al Quran dengan metode K-Means Clustering
--	---------------------	--	--	---

Kelebihan penelitian sekarang

- Sistem clustering data mahasiswa dibuat berdasarkan data mahasiswa baru sesuai dengan asal daerah masing-masing.
- Sistem bisa diakses melalui website sehingga mempermudah user.