

BAB II

LANDASAN TEORI

2.1 Pengertian Sistem

Menurut Tata Sutabri (2012:6) pada buku *Analisis Sistem Informasi*, pada dasarnya “sistem adalah sekelompok unsur yang erat hubungannya satu dengan yang lain, yang berfungsi bersama-sama untuk mencapai tujuan tertentu”.

Menurut McLeod (2004) dikutip oleh Yakub dalam buku *Pengantar Sistem Informasi* (2012:1) mendefinisikan “sistem adalah Sekelompok elemen-elemen yang terintegrasi dengan tujuan yang sama untuk mencapai tujuan. Sistem juga merupakan suatu jaringan kerja dari prosedur-prosedur yang saling berhubungan, terkumpul bersama-sama untuk melakukan suatu kegiatan atau untuk tujuan tertentu”.

2.1.1 Pengertian Informasi

Menurut McLeod dikutip oleh Yakub (2012:8) pada buku *Pengertian Sistem Informasi*, “Informasi adalah data yang diolah menjadi bentuk lebih berguna dan lebih berarti bagi yang menerimanya”.

Menurut Tata Sutabri (2012:22) pada buku *Analisis Sistem Informasi*, “Informasi adalah data yang telah diklasifikasikan atau diolah atau diinterpretasikan untuk digunakan dalam proses pengambilan keputusan”.

2.1.2 Pengertian Sistem Informasi

Menurut O’Brian dikutip oleh Yakub (2012:17) pada buku *Pengantar Sistem Informasi*, “sistem informasi (*information system*)

merupakan kombinasi teratur dari orang-orang, perangkat keras, perangkat lunak, jaringan komunikasi, dan sumber daya data yang mengumpulkan, mengubah, menyebarkan informasi dalam sebuah organisasi”.

2.2 Pengertian Arsip

Arsip (record) yang dalam istilah bahasa Indonesia ada yang menyebutkan sebagai “warkat”, pada pokoknya dapat diberikan pengertian sebagai: setiap catatan tertulis baik dalam bentuk gambar ataupun bagan yang memuat keterangan-keterangan mengenai sesuatu subyek (pokok persoalan) ataupun peristiwa yang dibuat orang untuk membantu daya ingatan orang (itu) pula”.

Atas dasar pengertian diatas, maka yang termasuk dalam pengertian arsip itu misalnya: surat-surat, kwitansi, faktur, pembukuan, daftar gaji, daftar harga, kartu penduduk, bagan organisasi, foto-foto dan lain sebagainya.

Berdasarkan Undang-Undang Nomor 7 Tahun 1971 tentang Ketentuan-Ketentuan Pokok Kearsipan, pasal 1 ayat a dan ayat b, menetapkan bahwa yang dimaksud dengan arsip adalah:

- a Naskah-naskah yang dibuat dan diterima oleh Lembaga-Lembaga Negara dan Badan-Badan Pemerintahan dalam bentuk corak apapun, baik dalam keadaan tunggal maupun berkelompok, dalam rangka pelaksanaan kegiatan pemerintahan.
- b Naskah-naskah yng dibuat dan diterima oleh Badan-Badan Swasta dan atau perorangan, dalam bentuk corak apa pun, baik dalam keadaan

tunggal maupun berkelompok, dalam rangka pelaksanaan kehidupan kebangsaan.

Selain dari pengertian di atas, arsip dapat diartikan pula sebagai suatu badan (agency) yang melakukan segala kegiatan pencatatan penanganan, penyimpanan dan pemeliharaan surat-surat/warkat-warkat yang mempunyai arti penting baik ke dalam maupun ke luar; baik yang menyangkut soal-soal pemerintahan maupun non-pemerintahan, dengan menerapkan kebijaksanaan dan sistem tertentu yang dapat dipertanggungjawabkan.

2.3 Pengertian Kearsipan

Menurut Barthos (2007), kearsipan adalah kegiatan pencatatan, penanganan, penyimpanan, dan pemeliharaan surat-surat atau warkat-warkat yang mempunyai arti penting baik ke dalam maupun ke luar dengan menerapkan kebijaksanaan dan sistem tertentu yang dapat dipertanggungjawabkan.

2.4 Algoritma SOM

Algoritma *Self Organizing Map* (SOM) merupakan suatu metode JST yang diperkenalkan oleh Professor Teuvo Kohonen pada tahun 1980an. SOM merupakan salah satu bentuk topologi dari *Unsupervised Artificial Neural Network* (*Unsupervised ANN*) dimana dalam proses pelatihannya tidak memerlukan pengawasan (target keluaran). SOM digunakan untuk mengelompokkan (*clustering*) data berdasarkan karakteristik/fitur-fitur data. (Shieh dan Liao, 2012).

Self Organizing Map (SOM) menyediakan suatu teknik visualisasi data yang membantu memahami data yang memiliki dimensi yang kompleks dengan mengurangi dimensi data kedalam peta. SOM juga merupakan konsep *clustering* dengan mengelompokkan data yang memiliki kemiripan tertentu. Oleh karena itu dapat dikatakan bahwa SOM mengurangi dimensi data dan menampilkan kesamaan antar data. Penggunaan SOM dalam penganalisaan data dapat mengakomodir data yang beragam mulai dari data dengan *single-dimensional* maupun *multi-dimensional* atau data dengan *singlevariat* maupun data yang bersifat *multi-variat*. Dalam data mining SOM lazim digunakan untuk teknik data mining yang berkaitan dengan pengelompokan data (*clustering*) dengan demikian, maka seringkali SOM yang digunakan untuk proses *clustering* disebut dengan metode SOM *clustering*.

2.5 *Feature Selection*

Tahap seleksi fitur (*feature selection*) bertujuan untuk mengurangi dimensi dari suatu kumpulan teks, atau dengan kata lain menghapus kata-kata yang dianggap tidak penting atau tidak menggambarkan isi dokumen sehingga proses pengklasifikasian lebih efektif dan akurat (Feldman & Sanger, 2007., Berry & Kogan 2010). Pada tahap ini tindakan yang dilakukan adalah menghilangkan *stopword* (*stopword removal*) dan *stemming* terhadap kata yang berimbuhan (Berry & Kogan 2010., Feldman & Sanger 2007).

Stopword adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen. Misalnya “di”, “oleh”, “pada”, “sebuah”, “karena”

dan lain sebagainya. Sebelum proses *stopword removal* dilakukan, harus dibuat daftar *stopword* (*stoplist*). Jika termasuk di dalam *stoplist* maka kata-kata tersebut akan dihapus dari deskripsi sehingga kata-kata yang tersisa di dalam deskripsi dianggap sebagai kata-kata yang mencirikan isi dari suatu dokumen atau *keywords*.

Setelah melalui proses *stopword removal* tindakan selanjutnya adalah yaitu proses *stemming*. *Stemming* adalah proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya (*stem*). Tujuan dari proses *stemming* adalah menghilangkan imbuhan-imbuhan baik itu berupa prefiks, sufiks, maupun konfiks yang ada pada setiap kata. Jika imbuhan tersebut tidak dihilangkan maka setiap satu kata dasar akan disimpan dengan berbagai macam bentuk yang berbeda sesuai dengan imbuhan yang melekatinya sehingga hal tersebut akan menambah beban *database*. Hal ini sangat berbeda jika menghilangkan imbuhan-imbuhan yang melekat dari setiap kata dasar, maka satu kata dasar akan disimpan sekali walaupun mungkin kata dasar tersebut pada sumber data sudah berubah dari bentuk aslinya dan mendapatkan berbagai macam imbuhan. Karena bahasa Indonesia mempunyai aturan morfologi maka proses *stemming* harus berdasarkan aturan morfologi bahasa Indonesia.

Berdasarkan penelitian sebelumnya, ada beberapa algoritma *stemming* yang bisa digunakan untuk *stemming* bahasa Indonesia diantaranya algoritma *confix-stripping*, algoritma Porter *stemmer* bahasa Indonesia, algoritma Arifin dan Sutiono, dan Algoritma Idris. Dimana,

Algoritma *confix-stripping* adalah algoritma yang akurat dalam *stemming* bahasa.

2.6 *Term Weighting*

Pembobotan kemunculan kata dalam suatu dokumen digunakan untuk perhitungan tingkat kemiripan antar dokumen dengan *query* (Abror, 2011). Salah satu metode pembobotan yang sering digunakan adalah *TF-IDF* (*Terms Frequency – Inverse Document Frequency*). Metode ini merupakan metode untuk menghitung nilai atau bobot suatu kata (*term*) pada dokumen. Metode ini akan mengabaikan setiap kata-kata yang tergolong tidak penting. Oleh sebab itu, sebelum melakukan metode ini, proses *stemming* dan *stopword removal* harus dilakukan terlebih dahulu oleh system (Pradnyana, 2012).

Faktor lain yang diperhatikan dalam pemberian bobot adalah kalimat yang muncul pada sedikit dokumen harus dipandang sebagai kata yang lebih penting daripada kalimat yang muncul pada banyak dokumen. Pembobotan akan memperhitungkan faktor kebalikan frekuensi dokumen yang mengandung suatu kalimat (*inverse document frequency*) (Pradnyana, 2012).

Nilai IDF sebuah *term* (kata) dapat dihitung menggunakan persamaan berikut:

$$\text{Idf}_t = \log \left(\frac{N}{\text{dft}} \right) \quad \text{Persamaan (1) untuk } N > \text{dft}$$

$$\text{Idf}_t = \log_2 \left(\frac{N}{\text{dft}} \right) \quad \text{Persamaan (2) untuk } N = \text{dft}$$

N adalah jumlah dokumen yang berisi *term* (t) dan n adalah jumlah kemunculan frekuensi *term* terhadap D. Adapun algoritma yang digunakan untuk menghitung bobot (W) masing-masing dokumen terhadap kata kunci (*query*) yaitu:

$$W_{dt} = tf_{d,t} \cdot idf_t \quad \text{Persamaan (3)}$$

Keterangan:

d= dokumen ke-d

t= *term* ke-t dari kata kunci

tf= *term* frekuensi

W= bobot dokumen ke-d terhadap *term* ke-t

Setelah bobot (W) masing-masing dokumen diketahui, maka dilakukan proses pengurutan (*sorting*) dimana semakin besar nilai W, maka semakin besar pula tingkat kesamaan (*similarity*) dokumen tersebut terhadap kata yang dicari, demikian pula sebaliknya.

2.7 Tinjauan Pustaka

Berikut ini beberapa penelitian tentang Algoritma SOM. Dalam penelitian “*A hardware configurable self-organizing map for real-time color quantization*” dalam karya yang disajikan, hardware dikonfigurasi Noc berbasis arsitektur SOM diterapkan untuk real-time image kuantisasi (Mehdi Abadi; Slavisa Jovanovic; Khaled Ben Khalifa; Serge Weber; Mohammed Hédi Bedoui, 2016). Dalam penelitian “*Animal call segregation using Self Organizing Map with Speeded Up Robust Features*” makalah ini menjelaskan titik awal untuk memisahkan audio spektrogram melalui penggunaan SOM yang dimodifikasi menggunakan tujuan umum

fitur gambar U-SURF (Kevin Pelzers, Valeri Mladenov, 2016). Dalam penelitian “*Automated Left Ventricle Posterior Wall Segmentation Using Kohonen Self-Organizing Map*” makalah ini mencoba untuk menemukan penerapan teknik SOM dalam membangun gambar 2D ekokardiografi dengan sinyal lemah dan resolusi spasial rendah (Salety, Vinicius, and Regina, 2016). Dalam penelitian “*Clustering and Classification Using a Self-Organizing Map*” Self-Organizing Map adalah jaringan saraf model yang membuang banyak keuntungan dan menyajikan kinerja yang sangat baik dalam proses klasifikasi. Namun, di beberapa kondisi tertentu, pertunjukan ini menghadapi beberapa kesulitan. Misalnya, dalam normalisasi ruang input, klasifikasi kehilangan presisi dan neuron tidak bisa membedakan antara input asli (Rafik Lasri, 2016). Dalam penelitian “*Comparative Study of Self-Organizing Map and Deep Self-Organizing Map using MATLAB*” mendalami Self-Organizing Map untuk perhitungan yang terdiri dari berbagai lapisan yang juga menggabungkan produksi peta 2D (Indra Kumar D and Manjunath R Kounte, Member, IEEE, 2016). Dalam penelitian “*Data Classification combining Self-Organizing Maps and Informative Nearest Neighbor*” karya ini akan menyajikan classifier data proposal yang menggabungkan SOM (Self-Organizing Map) dan saraf jaringan dengan INN (*Informative Nearest Neighbor*). Klasifikasi akan dilakukan dalam dua langkah: pertama SOM memiliki fungsi untuk mengurangi dataset melalui proses yang memanfaatkan konsep neuron dan pada langkah kedua, objek dataset yang dipilih akan dimanfaatkan sebagai referensi untuk algoritma INN

untuk memutuskan tentang klasifikasi memanfaatkan objek yang paling informatif dari dataset yang dikurangi (Leandro J. Moreira, Leandro A. Silva, 2016). Dalam penelitian “*Financial Self-Organizing Map of Croatian small-sized enterprises*” Penelitian telah mencoba pendekatan yang berbeda untuk analisis keuangan usaha kecil sektor kroasia serta untuk memilih rasio algoritma keuangan, cocok untuk pengembangan model clustering (Mario Zupan, Svjetlana Letinić, and Verica Budimir, 2016). Dalam penelitian “*Self-Organizing Map for Data Collection Planning in Persistent Monitoring with Spatial Correlations*” Dalam tulisan ini, generalisasi dari PC-tspn untuk data perencanaan koleksi dengan korelasi spasial antara pengukuran sensor telah diusulkan bersama dengan ekstensi dari algoritma berbasis SOM untuk mengatasi masalah umum (Jan Faigl and Petr Vana, 2016). Dalam penelitian “*Self-Organizing Maps for Multi-System and Multi-View Datasets*” tujuan dari penelitian ini adalah untuk mengembangkan skema meta-learning sebagai dasar kecerdasan buatan yang dapat mengidentifikasi meta-pengetahuan. Kami percaya bahwa SOSM dan SOVM mewakili titik awal yang baik untuk pengembangan mesin pembelajaran yang dapat menemukan meta-pengetahuan tentang mesin pembelajaran diri (Hideaki Ishibashi and Tetsuo Furukawa, 2016). Dalam penelitian “*Self-Organizing Map-based Solution for the Orienteering Problem with Neighborhoods*” dalam tulisan ini, mengusulkan algoritma berbasis SOM Novel untuk masalah Orienteering. Meskipun algoritma yang diusulkan tidak memberikan hasil yang lebih baik untuk semua masalah yang dievaluasi, yang utama

manfaatnya adalah kemampuan untuk memecahkan OP umum dengan jangkauan komunikasi non-nol, dilambangkan sebagai Orienteering yang berhubungan dengan masalah lingkungan (OPN) (Jan Faigl, Robert Penicka, and Graeme Best, 2016).

Berikut ini beberapa penelitian tentang metode TF-IDF. Dalam penelitian "*TF-SIDF: Term Frequency, Sketched Inverse Document Frequency*" dalam karya ini, telah mengusulkan TS-SIDF, sebuah solusi baru untuk mengatasi kompleksitas ruang yang terlibat dalam penggunaan jumlah yang tepat dalam perhitungan TF-IDF dengan memakai teknik aliran data, khususnya membuat sketsa algoritma (Manuel Baena-Garcia, Jose M. Carmona-Cejudo, Gladys Castillo and Rafael Morales-Bueno, 2011). Dalam penelitian "*Subjectivity Classification of Filipino Text with Features based on Term Frequency – Inverse Document Frequency*" Dalam tulisan ini menjajaki mesin yang berbeda, belajar algoritma pada mengklasifikasikan Artikel dan kalimat sebagai subjektif atau objektif (Regalado, Ralph Vincent J., Chua, Jenina L., Co, Justin L., Tiam-Lee, Thomas James Z., 2013). Dalam penelitian "*Automated Document Classification for News Article in Bahasa Indonesia based on Term Frequency Inverse Document Frequency (TF-IDF) Approach*" TF-IDF memiliki ketergantungan tinggi pada leksikon (Ari Aulia Hakim, Alva Erwin, Kho I Eng, Maulahikmah Galinium, Wahyu Muliady, 2014). Dalam penelitian "*Multi Words Quran and Hadith Searching Based on News Using TF-IDF*" dalam kasus ini aplikasi memberikan ketepatan 60% untuk ayat-ayat Alquran dan 53% untuk hadis. Evaluasi ahli memberikan

85,2% setuju dengan akurasi hasil ayat Quran, 85,1% setuju dengan kelengkapannya. 58,9% setuju dengan hasil hadis yang akurat, dan 61,8% setuju dengan kelengkapannya. Waktu eksekusi rata-rata adalah 2.706 detik (Eko Darwiyanto, Ganang Arief Pratama, Sri Widowati, 2016). Dalam penelitian “*Document Clustering: TF-IDF approach*” mengidentifikasi dokumen replika mendapat hasil yang setelah diolah dari berbagai dataset menunjukkan efisiensi dari algoritma. Menggunakan konsep relativitas semantik yang lebih baik akan menjadikan domain untuk memberikan hasil yang lebih baik (Prafulla Bafna, Dhanya Pramod, and Anagha Vaidya, 2016).

2.8 Perancangan Sistem

Penggambaran dan perancangan model sistem informasi secara grafik pada proses yang terjadi dalam sebuah alur.

a. *Flow Map*

Flowmap adalah penggambaran secara grafik dari langkah - langkah dan urutan urutan prosedur dari suatu program. *Flowmap* menolong analis dan programmer untuk memecahkan masalah kedalam segmen-segmen yang lebih kecil dan menolong dalam menganalisis alternatif - alternatif lain dalam pengoperasian. *Flowmap* biasanya mempermudah penyelesaian suatu masalah khususnya masalah yang perlu dipelajari dan dievaluasi lebih lanjut.

b. Diagram Konteks (*Context Diagram*)

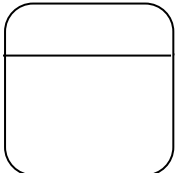

Diagram konteks adalah bagian dari data flow diagram yang berfungsi memetakan model lingkungan, yang dipresentasikan dengan lingkaran tunggal yang mewakili keseluruhan sistem. (Yakub:2012)



c. Diagram Alir Data (*Data Flow Diagram*)

Diagram Alir Data merupakan gambaran sistem yang lebih khusus dan lebih terinci dari pada contex diagram. Definisi menurut Tata Sutabri pada buku *Analisis Sistem Informasi* (2012:117), Data Flow Diagram adalah sebagai berikut : “*Data Flow Diagram* ini adalah suatu network yang menggambarkan suatu sistem automat/komputersasi, manualisasi, atau gabungan dari keduanya, yang penggambarannya disusun dalam bentuk kumpulan komponen sistem yang saling berhubungan sesuai dengan aturan mainnya.

Adapun simbol – simbol yang digunakan untuk menggambarkan *Data Flow Diagram* (DFD) terlihat pada Tabel 2.1:

Tabel 2.1 Simbol Data Flow Diagram (DFD)

No.	Simbol Data Flow Diagram	Keterangan
1.		Proses , kegiatan yang dilakukan oleh orang, mesin, atau komputer dari hasil suatu arus data yang dapat berupa masukan untuk sistem.
2.		External Entity , menyatakan kesatuan dilingkungan luar sistem yang dapat berupa orang, organisasi atau sistem yang lainnya yang berada di lingkungan luarnya yang akan memberikan input atau menerima output dari sistem.

3.		Simpanan Data , simpanan dari data yang dapat berupa suatu file atau data base di sistem komputer, suatu arsip atau catatan manual dan suatu agenda atau buku.
4.		Arus Data , menyatakan arus data yang mengalir diantara data proses, simpan data dan kesatuan luar. arus data ini menunjukkan arus dari data yang dapat berupa masukan untuk sistem atau hasil dari proses sistem.

Ada beberapa tahapan dalam pembuatan data flow diagram yaitu:

1. Diagram level konteks : digunakan untuk menggambarkan sistem secara global.
2. Diagram level nol : digunakan untuk menggambarkan tahapan-tahapan proses ada dalam diagram konteks.
3. Diagram detail (level) : digunakan untuk menggambarkan arus data yang lebih mendetail dalam proses diagram level nol.


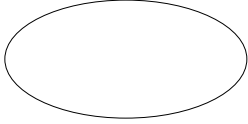
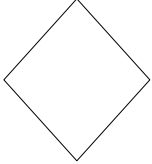

d. Kamus Data

Kamus data adalah suatu daftar data elemen yang terorganisir dengan definisi yang tetap dan sesuai dengan sistem, sehingga user dan analis sistem mempunyai pengertian yang sama tentang input, output, dan komponen data store. Kamus data ini sangat membantu analis sistem dalam mendefinisikan data yang mengalir di dalam sistem, sehingga pendefinisian data itu dapat dilakukan dengan lengkap dan terstruktur. Pembentukan kamus data dilaksanakan dalam tahap analisis dan perancangan suatu sistem. (Yakub:2012)

e. ERD (*Entity Relationship Diagram*)

Ladjamudin (2013:142), ERD adalah suatu model jaringan yang menggunakan susunan data yang disimpan dalam system secara abstrak. Notasi-notasi simbolik di dalam Diagram E-R yang dapat kita gunakan terlihat pada Tabel 2.2:

Tabel 2.2 Simbol ERD

No	Simbol	Keterangan
1.	Himpunan Entitas/ <i>Entity</i> 	Persegi panjang, menyatakan himpunan entitas.
2.	Atribut 	Lingkaran/Elip, menyatakan atribut (atribut yang berfungsi sebagai key digarisbawahi).
3.	Himpunan Relasi 	Belah ketupat, menyatakan himpunan relasi/relasi.
4.	Link 	Garis, sebagai penghubung antara himpunan relasi dengan himpunan entitas dan himpunan entitas dengan atributnya.

2.11 Teori Program

a. PHP

Ardhana (2012:88), PHP merupakan bahasa pemrograman berbasis server- side yang dapat melakukan parsing script php menjadi script

web sehingga dari sisi client menghasilkan suatu tampilan yang menarik

b. MySQL

Bunafit (2013:26), MySQL adalah software atau program database server.

c. PHP MyAdmin

Bunafit (2013:15), PHPMyAdmin adalah aplikasi manajemen database server MySQL berbasis web. Dengan aplikasi phpMyAdmin kita bias mengelola database sebagai root atau juga sebagai user biasa, kita bias membuat database baru, megelola database dan melakukan operasi perintah-perintah database secara lengkap seperti saat kita di MySQL Promp.