BAB 2
LANDASAN TEORI

2.1 Artikel

Artikel adalah tulisan lepas berisi opini seseorang yang mengupas tuntas suatu masalah tertentu yang sifatnya aktual dan atau kontroversial dengan tujuan untuk memberitahu (informatif), memengaruhi dan meyakinkan (persuatif argumentatif), atau menghibur khalayak pembaca (rekreatif). Disebut lepas, karena siapa pun pembaca boleh menulis artikel dengan topik bebas sesuai dengan minat dan keahliannya masing-masing. Selain itu juga artikel yang ditulis tersebut tidak terkait dengan berita atau laporan tertentu. Ditulisnya pun boleh kapan saja, di mana saja, dan oleh siapa saja. Secara teknis jurnalistik, artikel adalah salah satu bentuk opini yang terdapat dalam surat kabar atau majalah. Disebut salah satu, karena masih ada bentuk opini yang lain. Analoginya sederhana, kalau kita membuka halaman demi halaman surat kabar atau majalah maka secara umum isinya dapat digolongkan ke dalam tiga kelompok besar. Kelompok pertama adalah berita (news), kelompok kedua disebut opini (views), kelompok ketiga dinamakan iklan (advertising)[4].

Kelompok berita, meliputi berita langsung (*straight news*), berita foto (*photo news*), berita suasana-berwarna (*colour news*), berita menyeluruh (*comprehensive news*) berita mendalam (*depth news*), berita penafsiran (*interpretative news*), dan berita penyelidikan (*investigative news*). Kelompok opini, meliputi tajuk rencana atau editoril, karikatur, pojok, artikel, kolom, dan surat pembaca. Untuk memisahkan secara tegas antara berita (*news*) dan opini (*views*), maka tajuk rencana, karikatur, pojok artikel, dan surat pembaca ditempatkan pada satu halaman khusus. Pemisahan secara tegas berita dan opini tersebut merupakan konsekuensi dari norma dan etika luhur jurnalistik yang tidak menghendaki berita sebagai fakta objektif, diwarnai atau dibaurkan dengan opini sebagai pandangan yang bersifat subjektif [4].

2.2 Kalimat

Kalimat adalah satuan bahasa terkecil dalam wujud lisan atau tulisan, yang mengungkapkan pikiran yang utuh. Kalimat terdiri atas deret kata yang dimulai dengan huruf kapital dan diakhiri dengan tanda titik. Unsur-unsur kalimat terdiri dari kata, kelompok kata dan lagu kalimat. Di dalam kalimat terdapat pengaturan hubungan kedudukan antara bagian-bagiannya. Ada bagian didalam kalimat yang menunjukkan sebagai "pelaku", ada bagian yang menunjukkan sebagai "perbuatan", ada bagian yang menunjukkan "bagaimana perbuatan itu dilakukan". Berdasarkan jabatannya kalimat terdiri dari.

1. Subyek, yaitu bagian yang menjadi pangkal atau pokok pembicaraan.
2. Predikat, yaitu bagian yang menerangkan subyek, biasanya berdiri sesudah subyek.
3. Obyek, yaitu bagian yang menjadi tujuan.
4. Keterangan, yaitu bagian yang menunjukkan waktu (keterangan waktu), tempat (keterangan tempat), alat (keterangan alat) dan sebagainya. Sedangkan kalimat berdasarkan fungsinya, dapat dikategorikan sebagai berikut.
    a. Kalimat pernyataan.
    b. Kalimat pertanyaan.
    c. Kalimat perintah.
    d. Kalimat seruan.

2.3 Kata

Kata adalah kesatuan terkecil yang diperoleh sesudah kalimat dibagi atas bagianbagiannya dan mengandung suatu ide. Kategori kata berdasarkan sintaksisnya terdiri dari lima kata [7] yaitu :

1. Kata Benda (*Nomina*) Kata benda adalah kata yang mengacu pada manusia, binatang, benda dan konsep atau pengertian.
2. Kata Kerja (*Verba*) Kata kerja adalah kata yang menyatakan tindakan.

3. Kata Sifat (*Adjektiva*) Kata sifat adalah kata yang memberi keterangan yang lebih khusus tentang sesuatu yang dinyatakan oleh nomina dalam kalimat.
4. Kata Keterangan (*adverbia*) Kata keterangan adalah kategori yang dapat mendampingi adjektiva, numeralia atau preposisi dalam konstruksi sintaksis.
5. Kata tugas adalaah kata yang hanya memiliki arti gramatikal dan tidak memiliki arti leksikal.

2.4 Peringkasan Teks

Konsep sederhana ringkasan adalah mengambil bagian penting dari keseluruhan isi dari artikel. Ringkasan adalah mengambil isi yang paling penting dari sumber informasi yang kemudian menyajikan kembali dalam bentuk yang lebih ringkas bagi penggunanya[5].

2.4.1 Tahapan Membuat Ringkasan

Ada beberapa tahapan dalam membuat ringkasan [6], yaitu sebagai berikut.

1. Membaca naskah asli secara menyeluruh untuk mengetahui kesan umum, gagasan pengarang dan sudut pandangnya.
2. Mencatat semua gagasan u utama atau gagasan penting.
3. Menyusun kembali suatu karangan singkat berdasarkan gagasan tersebut.

Untuk membuat dan mendapatkan ringkasan yang baik, yaitu dengan cara membaca dengan seksama teks bacaan dan mencatat hal-hal pokok dari bacaan tersebut[6]. Ciri-ciri sebuah ringkasan yang baik itu harus mudah dipahami, bentuknya lebih singkat atau lebih ringkas, dan terdiri dari gagasan- gagasan utama (*expert judgement*).

2.5 *Text Mining*

Text mining adalah proses mencari informasi dengan menggunakan *tools* analisis berupa kategorisasi. Proses *text mining* dapat menjadi solusi dari permasalah pemrosesan data berupa teks yang tidak terstruktur dalam jumlah yang banyak. *Text mining* sering digunakan dalam kasus *information retrieval*. Tujuan adanya *text mining* yaitu untuk mendapatkan informasi yang bermanfaat dari sekumpulan data, salah satunya adalah dokumen teks [17].

2.6 Peringkasan Teks Otomatis

Peringkasan teks otomatis adalah teknik pembuatan ringkasan dari sebuah teks secara otomatis dengan memanfaatkan aplikasi yang dijalankan pada komputer untuk menghasilkan informasi yang paling penting dari dokumen aslinya [5].

2.6.1 Pendekatan Peringkasan Teks Otomatis

Terdapat 2 pendekatan peringkas teks[5], yaitu:

1. Ekstraksi(*extractive summari*). Pada teknik ekstraksi, sistem manyalin unit-unit teks yang dianggap paling penting dari sumber teks menjadi ringkasan. Unit-unit teks yang disalin dapat berupa klausa utama, kalimat utama, atau tidak terdapat pada paragraf utama tanpa ada penambahan-penambahan kalimat baru yang tidak teradapat pada dokumen aslinya.
2. Abstraksi (*abstraksi summary*). Teknik abstraksi menggunakan metode *linguistik* untuk memeriksa dan menafsirkan teks dokumen menjadi ringkasan. Ringkasan teks tersebut dihasilkan dengan cara menambahkan kalimat-kalimat yang ada pada teks sumber.

2.6.2 *Preprocessing*

Pada tahap pra proses (*preprocessing*) dilakukan penyiapan dokumen mentah menjadi dokumen atau representatif dokumen yang siap diproses untuk langkah selanjutnya[5]. Pada tahap ini proses yang dilakukan antara lain membagi dokumen

menjadi kalimat, *casefolding*, *filtering*, dokumen menjadi kata (*tokenizing*), dan menghapus *stopword*.

1. Pemisahan Kalimat.

Memecah dokumen menjadi kalimat-kalimat merupakan langkah awal tahapan preprocessing. Pemecahan kalimat yaitu proses memecah string teks dokumen yang panjang menjadi kumpulan kalimat-kalimat. Dalam memecah dokumen menjadi kalimat-kalimat menggunakan fungsi split (), dengan tanda titik (“.”) sebagai pemisah (delimiter) untuk memotong string dokumen[7].

2. *Case Folding*

Case folding adalah tahapan proses mengubah semua huruf dalam teks dokumen menjadi huruf kecil, serta menghilangkan karakter selain a-z. [7].

3. *Filtering*

Data teks dalam dokumen yang sebelumnya sudah diubah ke dalam huruf kecil semua. Selanjutnya dilakukan proses *filtering* teks. *Filtering* adalah tahapan pemrosesan teks dimana semua teks selain karakter “a” sampai “z” dan titik “.” akan dihilangkan dan hanya menerima spasi[7].

4. *Tokenizing*

Tokenizing adalah proses pemotongan string input berdasarkan tiap kata yang menyusunnya. Pemecahan kalimat menjadi kata-kata tunggal dilakukan dengan men-scan kalimat dengan pemisah (*delimiter*) *white space* (*spasi*, *tab*, dan *newline*)[7].

5. *Stopword*

Penghapusan Stopword merupakan proses penghilangan kata stopword. Stopword adalah kata - kata yang sering kali muncul dalam dokumen namun arti dari kata-kata tersebut tidak deskriptif dan tidak memiliki keterkaitan dengan tema tertentu. Misalnya “di”, ”oleh”, “pada”, ”sebuah”, ”karena” dan lain sebagainya[7].

2.7 Metode TF-IDF (*Term Frequency – Inversed Document Frequency*)

Metode *Term Frequency - Inverse Document Frequency (TF-IDF)* merupakan suatu cara untuk memperoleh pembobotan berdasarkan jumlah II-8 kemunculan suatu kata (*term*) dalam sebuah dokumen *term frequency (TF)* dan jumlah kemunculan term dalam koleksi dokumen *inverse document frequency (IDF).*

TF merupakan banyak kata yang muncul pada sebuah dokumen, sedangkan DF merupakan banyaknya dokumen yang mengandung sebuah kata. Nilai IDF sebuah kata (*term*) dapat dihitung menggunakan persamaan berikut.

$$\text{IDF} = log\ \frac{N}{df} \tag{2.1}$$

N adalah jumlah dokumen yang berisi *term* (t) dan df adalah jumlah kemunculan kata (*term*) terhadap N. Adapun rumus yang digunakan untuk menghitung bobot (W) masing-masing dokumen, yaitu dapat dilihat pada rumus berikut.

$$W_{dt} = tf_{dt} * \text{IDF}_t \tag{2.2}$$

dengan:

d = kalimat ke–d

t = kata (*term*) ke–t

TF = *term frequency* / frekuensi kata

W = bobot kalimat ke–d terhadap kata (*term*) ke–t

IDF = *Inverse Document Frequency*

Berikut ilustrasi dari TF-IDF:

dokumen 1 (d1) : Manajemen transaksi logistik.

dokumen 2 (d2) : Pengetahuan antar individu.

dokumen 3 (d3) : Dalam manajemen pengetahuan terdapat transfer pengetahuan logistik. Jadi jumlah dokumen (d) = 3

Setelah melalui proses filtering, maka kata “antar” pada dokumen 2 serta kata “dalam” dan “terdapat” pada dokumen 3 dihapus.

Tabel 2. 1 Contoh perhitungan TF-IDF

Token	Tf			DF	IDF	W (TF-IDF)		
	D1	D2	D3			D1	D2	D3
Manajemen	1	0	1	2	0.1761	0.1761	0	0.1761
transaksi	1	0	0	1	0.4771	0.4771	0	0
Logistik	1	0	1	2	0.1761	0.1761	0	0.1761
Pengetahuan	0	1	1	2	0.1761	0	0.1761	0.3522
Individu	0	1	0	1	0.4771	0	0.4771	0
transfer	0	0	1	1	0.4771	0	0	0.4771

Dari tabel di atas didapat :

Bobot (w) untuk d1 = 0.8293

Bobot (w) untuk d2 = 0.6532

Bobot (w) untuk d3 = 1.1815

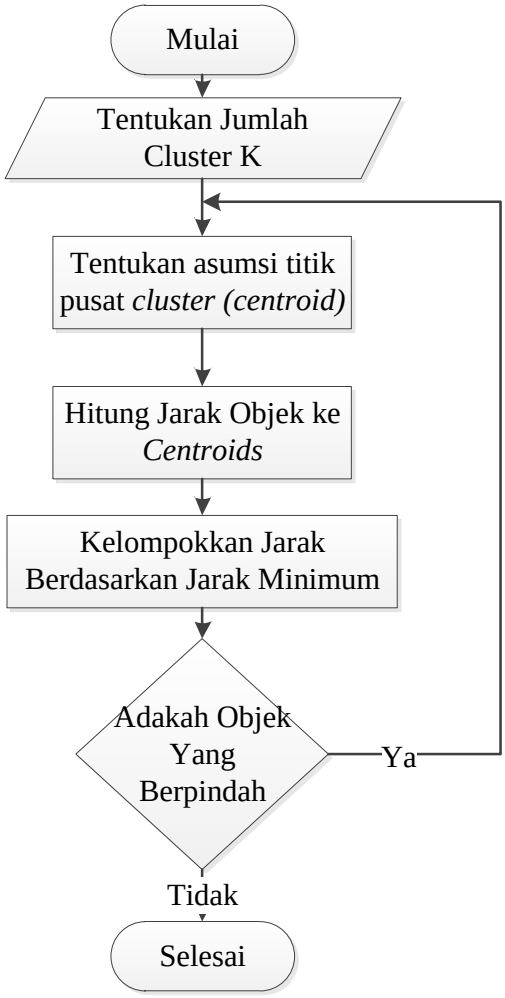
2.8 Metode *K-Means Clustering*

K-Means Clustering merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok[8].

Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan objective function yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu cluster dan

memaksimalisasikan variasi antar cluster. Manfaat *Clustering* adalah sebagai *Identifikasi Object (Recognition*) misalnya dalam bidang *Image Processing*, *Computer Vision* atau *robot vision*. Selain itu adalah sebagai Sistem Pendukung Keputusan dan Data Mining seperti Segmentasi pasar, pemetaan wilayah, Manajemen marketing dll.

Data *clustering* menggunakan metode *K-Means* ini secara umum dilakukan dengan algoritma dasar[9]. Algoritma K-means dapat dilihat pada gambar 2.1 berikut.



Gambar 2. 1 Algoritma *K-Means*

1. Tentukan jumlah *cluster*

2. Menentukan nilai *centroid*

Dalam menentukan nilai *centroid* untuk awal iterasi, nilai awal *centroid* dilakukan secara acak. Sedangkan jika menentukan nilai *centroid* yang merupakan tahap dari iterasi dihitung berdasarkan nilai rata-rata dari data yang terletak pada *centroid* yang sama.

3. Menghitung jarak antara data dengan pusat *cluster*.

Untuk menghitung jarak tersebut dapat menggunakan *Euclidean Distance. Euclidean* sering digunakan karena penghitungan jarak dalam *distance space* ini merupakan jarak terpendek yang bisa didapatkan antara dua titik yang diperhitungkan[8]. Berikut persamaan dengan *Euclidean Distance*.

$$De = \sqrt{(xi - yi)^2} \tag{2.3}$$

dengan:

D*e* = *euclidean distance.*

i = banyaknya data.

x = bobot dokumen.

y = pusat cluster.

4. Pengelompokan Data

Untuk menentukan anggota *cluster* adalah dengan memperhitungkan jarak terpendek data. Nilai yang diperoleh dalam keanggotaan data pada *distance* matriks adalah 0 atau 1, dimana nilai 1 untuk data yang dialokasikan ke *cluster* dan nilai 0 untuk data yang dialokasikan ke cluster yang lain.

5. Kembali ke tahap 2, lakukan perulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota *cluster* tidak berpindah ke cluster lain.

2.9 *Generating*

Tahapan ini adalah pembangkitan atau pembentukan hasil akhir. Terdiri dari penggabungan frase, pencetakan kata atau frase dan pembangkitan kalimat. Metode *generating* yang digunakan yaitu *Topic list* yang merupakan teknik hasil akhir yang berisi kata-kata yang sering muncul atau penggabungan pengertian yang telah diinterpretasi [7].

2.10 Teknik Evaluasi Peringkasan Teks

Pengujian yang dilakukan dalam penelitian ini menggunakan pengujian akurasi. Standar pengukuran yang biasa digunakan dalam penelitian *text summarization* yaitu *recall*, *precision*, dan *f-measure*[19].

Recall adalah tingkat keberhasilan ringkasan. Perhitungan *recall* dapat dilihat pada persamaan sebagai berikut.

$$recall = \frac{correct}{correct + missed} \tag{2.4}$$

dimana

recall : tingkat keberhasilan

correct : jumlah kalimat yang berhasil di ekstrak sistem sesuai dengan kalimat yang diekstrak manusia

missed : jumlah kalimat yang diekstrak manusia tetapi tidak terdapat dalam kalimat yang diekstrak sistem

Precision adalah tingkat ketepatan hasil ringkasan. Perhitungan *precision* dapat dilihat pada persamaan berikut:

$$precision = \frac{correct}{correct + wrong} \tag{2.5}$$

dimana,

precision : tingkat ketepatan

correct : jumlah kalimat yang berhasil di ekstrak sistem sesuai dengan kalimat yang diekstrak manusia

wrong : jumlah kalimat yang diekstrak sistem tetapi tidak terdapat dalam kalimat yang diekstrak manusia

F-measure adalah gabungan antara *recall* dan *precision*. Perhitungan *f-measure* dapat dilihat pada persamaan berikut:

$$F - meansure = \frac{2\ x\ recall\ x\ precision}{recall + \ precision} \quad (2.6)$$

2.11 *Flowchart*

Flowchart adalah bagan (*chart*) yang menunjukkan alir (flow) di dalam program atau prosedur sistem secara logika. Bagan alir program (*program flowchart*) merupakan bagan alir yang mirip dengan bagan alir sistem, yaitu untuk menggambarkan prosedur di dalam sistem [7].

2.12 *Data flow diagram*(DFD)

Pengertian Data Flow Diagram (DFD) adalah Diagram yang menggunakan notasi simbol untuk menggambarkan arus data system [7]. DFD sering digunakan untuk menggambarkan suatu sistem yang telah ada atau sistem yang baru yang akan dikembangkan secara logika dan menjelaskan arus data dari mulai pemasukan sampai dengan keluaran data tingkatan diagram arus data mulai dari diagram konteks yang menjelaskan secara umum suatu system atau batasan system dari level 0 dikembangkan menjadi level 1 sampai system tergambarkan secara rinci. Gambaran ini tidak tergantung pada perangkat keras, perangkat lunak, struktur data atau organisasi file.

Ada beberapa simbol digunakan pada DFD untuk mewakili [10].

1. Kesatuan Luar (*External Entity*)

Kesatuan luar (*external entity*) merupakan kesatuan (*entity*) di lingkungan luar sistem yang dapat berupa orang, organisasi, atau sistem lain yang berada pada lingkungan luarnya yang memberikan input atau menerima output dari sistem.

2. Arus Data (*Data Flow*)

Arus Data (*data flow*) di DFD diberi simbol suatu panah. Arus data ini mengalir di antara proses, simpan data dan kesatuan luar. Arus data ini menunjukan arus dari data yang dapat berupa masukan untuk sistem atau hasil dari proses sistem.

3. Proses (*Process*)

Proses (*process*) menunjukan pada bagian yang mengubah input menjadi output, yaitu menunjukan bagaimana satu atau lebih input diubah menjadi beberapa output. Setiap proses mempunyai nama, nama dari proses ini menunjukan apa yang dikerjakan proses.

4. Simpanan Data (*Data Store*)

Data Store merupakan simpanan dari data yang dapat berupa suatu file atau database pada sistem komputer.

2.13 Perangkat Lunak Pendukung

Pada bagian ini akan dijelaskan tentang beberapa perangkat lunak dan bahasa pemrograman yang digunakan untuk mendukung dalam pengembangan aplikasi yang dibuat. Bahasa pemrograman yang digunakan diantaranya adalah php dan html. Selain itu terdapat perangkat lunak yang digunakan yaitu XAMPP, Macromedia Dreamweaver dan MySQL

2.13.1 Pengertian XAMPP

XAMPP adalah satu paket software web server yang terdiri dari Apache, Mysql, PHP dan phpMyadmin. Mengapa menggunakan XAMPP? Karena XAMPP sangat mudah penggunaanya, terutama bagi pemula. Proses instalasi XAMPP sangat mudah, karena tidak memerlukan konfigurasu Apache, PHP dan MySQL secara manual, XAMPP melakukan instalasi dan konfigurasi secara otomatis [11].

2.13.2 Pengertian PHP

PHP merupakan kependekan dari kata *Hypertext Prepocessor*. PHP tergolong sebagai perangkat lunak open source yang diataur dalam aturan *general purpose licences* (GPL).

Pemograman PHP sangat cocok dikembangkan dalam lingkungan web, karena PHP bisa dilekatkan pada script HTML atau sebaliknya. PHP dikhususkan untuk pengembangan web dinamis. Maksudnya, PHP mampu menghasilkan website yang secara terus menerus hasilnya bisa berubah-ubah sesuai dengan pola yang diberikan. Hal tersebut bergantung pada permintaan *client browser* yang digunakan. Pada umumnya, pembuatan web dinamis berhubungan erat dengan database sebagai sumber data yang akan ditampilkan.

PHP tergolong juga sebagai bahasa pemograman yang berbasis server (*server side cripting*). Ini berarti bahwa semua script PHP diletakkan diserver dan diterjemahkan oleh web server terlebih dahulu, kemudian hasil terjemahan itu dikirim ke browser client. Tetu hal tersebut berbeda dengan JavaScript selalu tampak dihalaman web bersangkutan, jika dilakukan penyimpanan terhadap file web. Secara teknologi, bahasa pemograman PHP memiliki kesamaan dengan bahasa ASP (*Active Server Page*), *Cold Fusion*, JSP (*Java Server Page*), ataupun Perl.

2.13.3 Sublimtext

Sublime Text merupakan editor HTML yang professional mendesain, menulis kode program dan mengembangkan *website*, halaman *web*, dan aplikasi *web*. Dalam pengerjaanya dreamweaver memberikan tiga pilihan yaitu bekerja dengan menulis

kode program (*Menu Code*), dengan pengeditan secara visual (*Mode Design*) dan dengan tampilan keduanya (*Split Mode*). Dreamweaver juga menyediakan alat-alat bantu untuk mengembangkan kreatifitas pembuatan *web*.