

# Perancangan Model Penilaian Kualitas Informasi di Fanpage Facebook

Asep Somantri

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pasundan  
Jln. Dr. Setiabudhi no. 193 Bandung, Jawa Barat  
[somantri@unpas.ac.id](mailto:somantri@unpas.ac.id)

**Abstrak** - Pada saat terjadinya pertumbuhan kuantitas ketersediaan informasi, maka kualitas informasi menjadi faktor yang sangat penting, agar dapat dimanfaatkan oleh organisasi maupun individu untuk menentukan sikap terhadap situasi yang sedang dihadapi. Banyak pengguna Facebook yang menjadikan informasi yang ditulis oleh pemilik Fanpage Facebook sebagai pijakan untuk menentukan sikap terhadap situasi yang terjadi pada lingkungannya. Salah satu hal penting yang perlu disadari oleh pengguna Facebook adalah tidak semua informasi yang ditulis oleh pemilik Fanpage merupakan informasi yang berkualitas, sehingga belum tentu dapat dijadikan pijakan untuk menentukan sikap terhadap situasi yang sedang dihadapi. Hal tersebut mendorong dilakukannya penelitian ini yang bertujuan untuk membuat sebuah model yang diharapkan dapat dimanfaatkan sebagai panduan untuk menilai informasi yang ditulis oleh sebuah Fanpage Facebook. Model dibentuk dengan menggunakan metode inovatif dengan tahapan yang disusun berdasarkan hasil kajian terhadap beberapa konsep dan teori dari penelitian terdahulu, yaitu konsep informasi pada media sosial, metode pengolah teks, teori kualitas informasi dan kredibilitas sumber data.

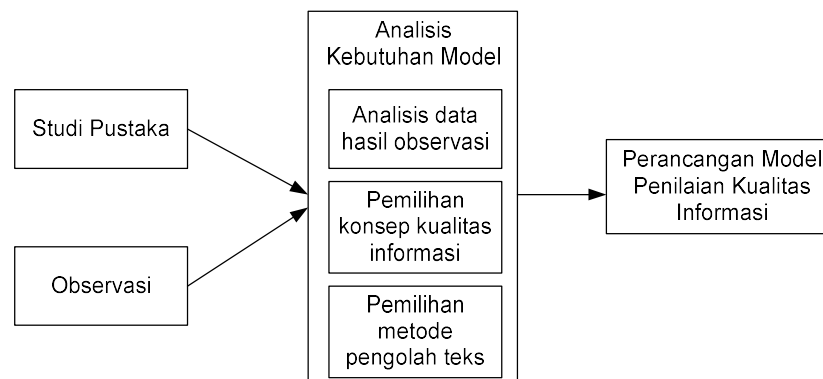
**Kata Kunci** - Fanpage Facebook, Kualitas Informasi, Model Penilaian, Perancangan, metode inovatif

## I. PENDAHULUAN

Terdapat beragam cara yang dapat ditempuh untuk menilai kualitas informasi di berbagai lingkungan. Beberapa model penilaian kualitas informasi-pun telah diusulkan oleh para peneliti sesuai dengan objek yang dinilainya. Sebagaimana telah dijelaskan oleh Kargar (2011) yang menyatakan bahwa, secara umum model penilaian kualitas informasi dibagi menjadi dua yaitu general purpose model dan special purpose model. Sosial media menghasilkan data yang termasuk kategori tidak terstruktur [2]. Beberapa peneliti membuat special purpose model untuk menilai data/ informasi yang tidak terstruktur, seperti SPADE [3], yaitu model yang dibuat untuk mendeteksi tulisan kebencian, spam, dan rumor dari media sosial yang berdampak kepada rendahnya kualitas informasi. Selain itu, ada model SNA [4], yaitu model yang dibuat untuk menyimpulkan seseorang berada pada komunitas tertentu berdasarkan hasil analisis interaksi data. Sebagaimana model SPADE dan SNA, penelitian ini ditujukan untuk membuat special purpose model, dan hal yang unip pada penelitian ini adalah model dibuat sebagai panduan untuk menilai kualitas informasi pada Fanpage Facebook.

## II. METODE PENELITIAN

Teori yang dikaji untuk digunakan dalam penelitian meliputi kualitas informasi, model penilaian, faktor-faktor penentu kualitas informasi di sosial media, dan metode pengolah teks yang relevan. Tahapan dimulai dengan melakukan studi pustaka untuk menemukan sekumpulan teori dan konsep yang dapat digunakan untuk merancang model penilaian kualitas informasi. Bersamaan dengan hal tersebut, dilakukan juga observasi terhadap perilaku dan kegiatan interaksi antar pengguna Facebook dan pengelola Fanpage Facebook.



Gambar 1 Metode Penelitian

Tahap analisis kebutuhan model dilakukan dengan tiga sub tahapan, yaitu analisis data hasil observasi yang akan menghasilkan Fanpage dan postingan terpilih untuk dinilai kualitas informasinya. Sub tahapan kedua adalah proses untuk memilih konsep dan teori kualitas informasi yang relevan untuk Fanpage Facebook dan sub tahapan ketiga adalah metode pengolahan teks yang cocok untuk mengolah informasi tidak terstruktur berupa teks. Tahapan terakhir adalah melakukan penyusunan setiap faktor yang telah dianalisis sebagai faktor yang dibutuhkan untuk menyusun model.

#### A. Fan Page Facebook

Jutaan bahkan milyaran orang dan organisasi menggunakan media sosial seperti Facebook, Twitter, YouTube, dan Google+ sebagai bagian dari kegiatan online harian [7]. Hal tersebut merupakan kegiatan pertukaran atau berbagi informasi yang sangat besar. *User* menggunakan sosial media untuk bersosialisasi dengan keluarga teman, saudara, rekan bisnis, dan kenalan. Meskipun jaringan sosial bertujuan untuk kegiatan professional, tetapi pada prakteknya tidak dominan, karena ada juga yang menggunakannya untuk hal lain seperti sekedar mencari rekan [8]

Fan Page Facebook adalah salah satu fitur yang disediakan oleh Facebook yang dapat digunakan sebagai sarana penyebaran informasi dari pemilik akun Fan Page ke beranda pemilik akun personal. Fan Page Facebook dapat dimiliki oleh perusahaan, organisasi, komunitas, maupun personal yang menginginkan dirinya digemari oleh khalayak ramai. Fan Page Facebook dapat digunakan untuk kepentingan bisnis, politik, penyebaran doktrin, pendidikan, kegiatan amal, dll.

#### B. Kualitas Informasi

Informasi adalah data yang telah diproses dan memiliki arti bagi penggunaannya [9]. Kualitas dapat menjadi konsep yang membingungkan, karena orang yang melihat kualitas menggunakan perspektif dan dimensi yang berbeda berdasarkan kepada peran individu masing-masing [10]. Dalam dunia penelitian, hal ini menunjukkan bahwa kriteria kualitas relatif terhadap perspektif dan dimensi masing-masing peneliti.

Kualitas informasi adalah ilmu untuk memaksimalkan nilai aset informasi dan menjamin informasi yang dihasilkan mampu memenuhi kebutuhan penggunaannya [11]. Kualitas informasi (*information quality*) lebih dikenali sebagai konsep dibandingkan variabel yang dapat diukur [11], sehingga dapat dipahami bahwa pembentukan model penilaian kualitas informasi dapat terus dikembangkan sesuai dengan kebutuhan peneliti.

#### C. Proses Pembentukan Model

Kargar (2011) menyatakan bahwa literatur yang ada telah mengusulkan sekian banyak model penilaian kualitas informasi yang berbeda, model ini mencakup tujuan umum (*general purpose model*) dan tujuan khusus (*special purpose model*). *General purpose model* adalah model yang memiliki kemampuan secara spesifik, tetapi tidak sama dengan model untuk tujuan khusus. Dengan kata lain, kriteria seleksi dan definisi untuk *special purpose model* tidak tergantung kepada lingkungan dan kerangka informasi. Tujuan dari *special purpose model* yang digunakan oleh setiap peneliti adalah untuk pencocokan model penilaian kualitas informasi terhadap aplikasinya masing-masing.

Terdapat beberapa contoh *general purpose model* yang populer seperti AIMQ [12], TDQM [13], dan Naumann [14]. Kargar (2011) menambahkan bahwa *special purpose model* berbeda dengan *general purpose model*. *Special purpose model* mengembangkan kriteria kualitas informasi berdasarkan kebutuhan masing-masing terhadap aplikasinya secara spesifik, seperti *Data Warehouse Quality* (DWQ), IQIP untuk perolehan informasi, aplikasi intranet, dan kualitas informasi di wikipedia.

Penelitian ini ditujukan untuk mengusulkan model penilaian kualitas informasi di sosial media. Berdasarkan penjelasan Kargar (2011), dapat dipahami bahwa model penilaian yang diperlukan adalah *special purpose model*.

#### D. Mengukur Kualitas Informasi

Fogg & Tseng (1999) menerangkan, hampir semua ahli kredibilitas menjelaskan bahwa kredibilitas sebagai kualitas yang dirasakan [15]. Adapun cara yang dapat ditempuh untuk mengatasi permasalahan kualitas informasi adalah menilai kualitas informasi berdasarkan kredibilitas sumber data. Oleh karena itu, model dibuat dengan berfokus kepada kredibilitas sumber data.

#### E. Kajian Faktor-faktor

Sosial media menyediakan sekian banyak data yang dibuat oleh user berupa tulisan-tulisan (*status update/ messages*), profil, dan *engagement*. Hal-hal tersebut dapat dijadikan material untuk dinilai kredibilitasnya. *Fan Page* pada *Facebook* menyediakan informasi mengenai profil dan konten (tulisan).

Penilaian yang dilakukan adalah dengan menetapkan nilai faktor perbedaan untuk menilai kredibilitas user berdasarkan keahliannya pada topik tertentu yang dibahasnya. Faktor-faktor yang dijadikan sebagai indikator untuk menilai kredibilitas sumber data diadopsi dari penelitian yang dilakukan oleh Canini et al. (2011)[16], lalu disesuaikan dengan kebutuhan penelitian. Berikut adalah faktor-faktor tersebut:

##### 1. Akun Fan Page

*Fan Page* atau halaman penggemar merupakan sebuah halaman seperti blog yang menyediakan informasi apapun sesuai dengan keinginan pemiliknya (seperti pendidikan, profil perusahaan, komunitas, artis, dakwah pengenalan produk dan layanan). Satu akun *Fan Page* dapat disukai atau diikuti oleh ribuan bahkan jutaan penggemar, sehingga *Fan Page* merupakan sarana yang cukup

efektif dalam menyampaikan informasi kepada penggemarnya. *Fan Page* yang dipilih untuk dinilai kredibilitasnya adalah *Fan Page* yang sering menyampaikan informasi penting yang berkaitan dengan kemaslahatan orang banyak.

## 2. Profil *Fan Page*

Perolehan informasi profil *Fan Page* merupakan salah satu yang penting dalam penilaian kredibilitas, karena profil dapat memuat informasi tentang pemilik *Fan Page* (seperti keahlian, pekerjaan, ketertarikan, karya-karya yang telah berhasil dibuat, prestasi yang pernah diraih dan pendidikan)

## 3. Status sosial

Canini et. al. (2011) menetapkan indikator profil pada twitter dalam beberapa hal, yaitu jumlah pengikut, jumlah akun yang diikuti, Jumlah tulisan yang dibuat, kemudian jumlah keanggotaan. Berdasarkan hal tersebut tersebut, dilakukan penyesuaian untuk *Fan Page* Facebook yang memiliki fitur yang serupa tapi tidak sama dengan Twitter. Adapun penyesuaian yang dilakukan adalah mempertimbangkan jumlah penggemar, postingan yang disukai, dikomentari dan dibagikan oleh penggemar.

## 4. Sekumpulan kata (*word clouds*).

Facebook memberikan fasilitas kepada user untuk membuat tulisan, penelitian ini memanfaatkan tulisan tersebut untuk dijadikan sebagai data yang dianalisis dengan menggunakan metode pengolahan kata untuk mengetahui topik yang sering dibahas.

## F. Metode Pengolah Kata

Tulisan yang dibuat dan disebar oleh pengguna sosial media berisi topik bahasan tertentu yang sesuai dengan ketertarikannya. Jika dikaitkan dengan pernyataan Fogg & Tseng (1999) tentang *expertise*, maka dapat disimpulkan bahwa tulisan pengguna sosial media mengindikasikan pengetahuan, pengalaman dan kompetensi yang dimiliki olehnya, sehingga topik yang sering dibahas oleh seseorang menjadi layak untuk dijadikan komponen dalam menilai kredibilitas dirinya. Sosial media menyediakan fasilitas kepada penggunanya untuk membuat tulisan sesuai dengan topik yang diinginkan. Setiap pengguna dapat membahas banyak topik, dan suatu topik dapat dibahas berulang kali, terlebih jika pengguna tertarik dengan topik tersebut. Kemudian, untuk penyimpulan topik memerlukan metode pengolahan kata.

## G. Persiapan Mengolah Kata

Terdapat suatu konsep yang dapat digunakan untuk mengolah setiap kata yang digunakan dalam sebuah tulisan, kemudian menyimpulkannya, sehingga ditemukan suatu topik tertentu yang mewakili seluruh tulisan, konsep tersebut dinamai LDA dan *Keyword Sentence Clustering*, namun sebelum hal tersebut dilakukan, maka diperlukan persiapan terlebih dahulu dengan tahapan sebagai berikut:

- 1) *Praproses*: Praproses merupakan tahapan awal sebelum dilakukannya pengolahan teks dengan menggunakan teknik-teknik untuk normalisasi teks. Normalisasi teks adalah strategi untuk meningkatkan ketepatan perolehan dokumen [20]. Beberapa teknik yang digunakan adalah *case folding*, *tokenisasi*, dan penghilangan *stopword*
- 2) *Case folding*: *Case folding* adalah salah satu teknik normalisasi teks untuk mengubah seluruh tulisan menjadi *lowercase* atau *uppercase* [21]. *Case folding* juga dilakukan untuk mengurangi *noise* pada dokumen, yaitu menghilangkan karakter berupa delimitter seperti titik (.), koma (,), titik dua (:), dan *underscore* (\_).
- 3) *Tokenisasi*: Tokenisasi adalah proses untuk memisahkan kata dari *body* teks dengan cara mengkonversi teks menjadi sekumpulan token, dan setiap token dinyatakan sebagai satu kata [21]. Terdapat banyak kata yang digunakan untuk membuat sebuah tulisan, kemudian setiap kata dapat digunakan berulang-ulang.
- 4) *Stopword Elimination*: Sebagaimana penelitian yang dilakukan oleh Blei (2003) mengabaikan kata yang masuk dalam daftar *stop word*. Demikian juga dengan penelitian ini. *Stop word* adalah kata umum yang tidak memberikan informasi penting [20]. Fakta yang ditemukan oleh Yates & Neto (1999) adalah kemunculan kata sebanyak 80 % pada dokumen tidak memiliki arti terhadap tujuan diambilnya informasi, sehingga keberadaannya diabaikan (lihat lampiran B). Kata-kata yang masuk dalam daftar *stopword* adalah preposisi dan konjungsi [20]. Preposisi adalah kata yang biasa terdapat di depan nomina, seperti “dari”, “dengan”, “di”, dan “ke”, sedangkan konjungsi adalah kata atau ungkapan penghubung antarkata, antarfrasa, antarklausa, dan antarkalimat, misalnya “dan”, “tetapi”, “atau” [22].

## H. Latent Dirichlet Allocation (LDA)

LDA (*Latent Dirichlet Allocation*) adalah model topik yang populer dan dapat digunakan untuk menemukan satu set topik berdasarkan sekumpulan kata dari tulisan yang dibuat [6]. Tulisan yang dibuat oleh seseorang dari sosial media, dalam hal ini disebut sebagai dokumen. Ide dasar dari LDA adalah menunjukkan sekian banyak kemungkinan nama topik untuk sebuah dokumen yang hendak dianalisis, dimana Sebuah topik dapat ditentukan/ disimpulkan berdasarkan sekumpulan kata yang digunakan [6].

## I. Penyimpulan Topik Dengan Cara Pengelompokan Kata

Terdapat tiga tahapan dasar yang digunakan untuk menyimpulkan topik dari sebuah dokumen [1]. Penelitian dilakukan untuk meningkatkan ketepatan ekstraksi kalimat yang representatif dan mengurangi susunan kalimat yang tidak sesuai dalam penyimpulan topik dari beberapa dokumen [13]. Terdapat lima tahapan untuk menyimpulkan topik dokumen yaitu, pertama menyiapkan dokumen yang hendak dianalisis. Kedua, mengelompokkan kalimat pada dokumen dengan menggunakan similarity histogram clustering (SHC). Ketiga, membuat peringkat cluster menggunakan cluster penting. Keempat, memilih kalimat yang representatif dari setiap *cluster* dengan mengidentifikasi kepadatan informasi dan kata kunci dari kelompok kalimat. Kelima, menyimpulkan kalimat representatif yang diperoleh dari setiap *cluster*.

Seluruh tahapan dilakukan dengan persamaan 1 hingga 11 [13]. Kemiripan kalimat berdasarkan pengelompokan histogram (SHC) adalah salah satu metode yang dapat memonitor dan menjaga keterkaitan *cluster* [5]. SHC dapat menjaga keterkaitan (*coherence*) dari setiap *cluster* dengan baik, hal ini ditunjukkan dengan pengurangan kalimat yang berlebih (*redundant*). Kemudian memonitor keterkaitan (*coherence*) melalui sejumlah komponen yang memiliki nilai kesamaan yang tinggi. Pengukuran yang digunakan untuk mengukur kemiripan antara dua kalimat adalah mencocokkan unigram berdasarkan pengukuran kemiripan dengan persamaan berikut.

$$sim(s_i, s_j) = \frac{(2 \times |s_i \cap s_j|)}{|s_i| + |s_j|} \quad (1)$$

$s_i$  dan  $s_j$  adalah kalimat ke- $i$  dan ke- $j$ .  $|s_i + s_j|$  adalah sejumlah kata yang cocok antara kalimat ke- $i$  dan ke- $j$ .  $|s_i|$  adalah panjang kalimat yang terdiri dari sejumlah kata.

*Histogram similarity* juga dapat menunjukkan kesesuaian cluster yang telah dibangun. Kesesuaian tersebut menunjukkan nilai yang sama antara satu kalimat dengan kalimat lain. Berikut ini adalah rumus yang digunakan untuk menghitung nilai tersebut.

$$H = count(sim_j) \text{ for } sim_{li} \leq sim_j \leq sim_{ui} \quad (2)$$

$sim_{li}$  adalah nilai kesamaan batas bawah dan  $sim_{ui}$  adalah nilai kesamaan batas atas.

Rasio histogram (HR) untuk sebuah cluster dihitung dengan menggunakan rumus berikut.

$$HR = \frac{\sum_{i=T}^{n_b} h_i}{\sum_{j=1}^{n_b} h_j} \quad (3)$$

$$T = |S_T \times n_b| \quad (4)$$

$S_T$  adalah batas nilai kemiripan (*similarity threshold*) yang ditunjukkan oleh  $T$ . Sebagaimana yang dijelaskan oleh IBM yang menyatakan bahwa *similarity threshold* adalah batas bawah untuk kesamaan dua data yang hendak dimasukkan ke dalam cluster yang sama [17].

Persamaan 1 hingga persamaan 4 digunakan pada model pengolah kata berdasarkan sekumpulan kata (*bag of word*). Representasi model pengolah kata berdasarkan sekumpulan kata tidak menangkap arti semantik dalam menerangkan dokumen, semantik hanya baik digunakan untuk menangkap kedekatan kata dan frekuensi kemunculannya dalam dokumen [19]. Penyimpulan topik dilakukan tidak berdasarkan sekumpulan kata, tetapi dilakukan berdasarkan sekumpulan frase (*bag of phrase*), karena representasi pengolah kata berdasarkan sekumpulan frase memiliki performansi lebih baik dalam menerangkan topik dokumen [19]. Setiap frase yang sama ditetapkan pada satu cluster yang sama, kemudian setiap *cluster* dinilai untuk memperoleh *cluster* yang penting.

```

Require: Word-topic matrix  $\Phi$  of dimension  $d \times k$ , List of
word-pairs  $L$  in the form of tuple  $\langle \omega_1, \omega_2 \rangle$ 
for  $j = 1$  to  $j = k$  do
    Sort column  $\Phi_{(j)}$  and create topic-list  $l_j$  of top  $t$  words
end for
for each list  $l_j$  from the set of topic-lists do
    for each word  $w_i \in l_j$  do
         $\omega_1 = w_i$ 
        if  $P(\omega_2 | z = j) \geq \tau$  where  $\omega_2$  is the corresponding
word of  $w_i$  from the word-pair list then
            add  $\langle \omega_1, \omega_2 \rangle$  to the list of phrases  $P$ 
        end if
    end for
end for
return list  $P$  of phrases generated
    
```

Gambar 2 Algoritma untuk Ekstraksi Frase [19]

Persamaan ke 1 dan ke 2 disesuaikan dengan kebutuhan penelitian, yaitu digunakan untuk menangkap frase yang sama dalam satu kalimat. Jumlah frase yang ditangkap adalah 2 frase hingga 5 frase. Penangkapan frase menggunakan persamaan ke 1 dan persamaan ke 2, kemudian dikonversikan dalam algoritma seperti terdapat pada gambar 1. *Cluster importance* digunakan untuk menyusun *cluster* yang terbentuk oleh SHC. *Cluster importance* menyusun pengelompokkan kalimat berdasarkan jumlah dari bobot frekuensi penggunaan kata yang ada dalam *cluster*. Jika terdapat  $N$  *cluster*, maka penyusunan dilakukan dengan menghitung bobot setiap *cluster* menggunakan rumus berikut.

$$Weight(C_j) = \sum_{w \in c_j} \log(1 + count(w)) \quad (5)$$

$C_j$  adalah *cluster* ke- $j$ ,  $count(w)$  adalah jumlah kata  $w$  di keseluruhan dokumen dan  $count(w)$  harus lebih tinggi dari batas nilai  $\theta$ . Bobot tersebut menunjukkan seberapa banyak informasi yang dimiliki *cluster* tersebut.

Kata kunci dari *cluster* kalimat (KCK) menggunakan pendekatan berdasarkan frekuensi kata yang digunakan dalam sebuah *cluster* dan juga distribusinya dalam *cluster* yang lain. KCK digunakan untuk menentukan bobot sebuah kata. Bobot kata yang bernilai tinggi merupakan kata yang representatif sebagai kata kunci dari sebuah *cluster*. Bobot setiap kata dihitung dengan menggunakan rumus berikut

$$iscf_{wij} = \log \frac{N}{scf_{wij}} \quad \text{dan} \quad (6)$$

$$tf \cdot iscf_{wij} = tf_{wi} Cluster_j \times iscf_{wij} \quad (7)$$

$tf_{wi} cluster_j$  adalah jumlah semua kata ke- $i$  dalam *cluster* ke- $j$ ,  $scf_{wij}$  adalah jumlah *cluster* yang mengandung kata  $w$  dalam *cluster* ke- $j$ , dan  $N$  adalah jumlah dari semua *cluster*. Nilai KCK dihitung dengan rumus berikut.

$$F_{kck}(s_{kj}) = \frac{1}{len(s_{kj})} \sum_{w_{ij} \in s_{kj}} tf \cdot iscf_{wij} \quad (8)$$

$F_{kck}(s_{kj})$  adalah nilai dari kalimat berdasarkan total bobot dari seluruh kata yang tersusun pada kalimat  $s$  ke- $k$  dalam *cluster* ke- $j$ ,  $len(s_{kj})$  adalah panjang kalimat  $s$  ke- $k$  dalam *cluster* ke- $j$ . Panjang dari kalimat ke- $k$  adalah jumlah kata yang menyusun kalimat tersebut.

Konsep penghitungan yang terdapat pada rumus ke 6 sampai 11 tidak digunakan sebagaimana penelitian Lukmana, dkk. (2014) yang menggunakannya untuk menemukan **kata kunci** dari setiap kalimat, tetapi dilakukan penyesuaian terhadap hasil penelitian Gujraniya, dkk. (2012), yaitu menemukan **topik kunci** dari keseluruhan dokumen.

Kalimat yang terpilih dari setiap *cluster* digunakan untuk membuat kesimpulan tentang topik yang menjadi bahasan sebuah dokumen. Kesimpulan tersebut dibuat berdasarkan kalimat yang dinilai representatif dan diperoleh dari *cluster* yang paling penting. Kalimat yang representatif dipilih secara berulang-ulang, sehingga kesimpulan topiknya sesuai dengan dokumen. Hal tersebut memunculkan hasil analisis yang menunjukkan topik yang gemar dibahas oleh pemilik akun sosial media. Dimana ketertarikan seseorang dalam membahas suatu hal mengindikasikan bahwa bidang keahliannya berada pada bidang yang sering dibahasnya tersebut.

#### J. Penyimpulan Topik berdasarkan Sekumpulan Kata

LDA menunjukkan metode untuk pemilihan kata kunci dari setiap dokumen. Lukmana (2014) memberikan bobot kepada kalimat berdasarkan kata kunci yang terpilih. Tahapan pertama, yaitu menggunakan setiap kata pada satu set dokumen menjadi bagian dari topik secara acak, kemudian topik yang terbentuk dapat dihitung kembali dengan menggunakan rumus

$$P(D) = \frac{D_T}{W_D} \quad (9)$$

$$P(D) = \frac{W_T}{N_T} \quad (10)$$

$P(D)$  adalah probabilitas setiap dokumen,  $D_T$  adalah jumlah dokumen yang memiliki topik  $T$ ,  $W_D$  adalah jumlah kata yang berada pada dokumen, kemudian  $P(T)$  adalah probabilitas topik-topik,  $W_T$  adalah jumlah kata pada topik  $T$  dan  $N_T$  adalah jumlah topik yang tersedia. Pembobotan kalimat dengan menggunakan LDA dihitung dengan rumus berikut

$$F_{LDA}(s_{kj}) = \frac{1}{len(s_{kj})} \sum_{w_{ij} \in s_{kj}} P_{wij} \quad (11)$$

$F_{LDA}$  adalah nilai dari kalimat berdasarkan total bobot dari seluruh kalimat yang menyusun kalimat ke- $k$  dalam *cluster* ke- $j$ , kemudian  $len(s_{kj})$  adalah panjang setiap kalimat ( $s$ ) ke- $k$  dalam *cluster* ke- $j$ . Panjang dari kalimat ke- $k$  adalah jumlah seluruh bobot dari kata yang menyusun kalimat tersebut. Kemudian,  $P_{wij}$  adalah probabilitas nilai dari kata ke- $i$  dalam *cluster* ke- $j$ .

#### K. Evaluasi Metode Penyimpulan Topik

Tidak mudah menilai baik atau buruknya suatu simpulan topik yang dihasilkan. Evaluasi yang dilakukan secara manual biasanya bersifat subjektif dan umumnya membutuhkan banyak usaha dari manusia [23], tetapi terdapat metode populer yang dapat digunakan untuk mengukur evaluasi simpulan topik secara otomatis, yaitu ROUGE (Recall Oriented Understudy for Gisting Evaluation). ROUGE mengandung langkah-langkah secara otomatis untuk menentukan kualitas simpulan topik dengan cara membandingkannya dengan 16 simpulan topik yang dibuat oleh manusia [24]. ROUGE merupakan metode evaluasi hasil ringkasan yang mengukur kualitas hasil ringkasan berdasarkan kesesuaian antara unit-unit ringkasan hasil sistem dengan unit-unit ringkasan referensi yang dibuat secara manual [23]. Penghitungan ROUGE-W yang diadopsi dari penghitungan Lin (2004) ditunjukkan oleh persamaan 12

$$R_{wlcs} = f^{-1} \left( \frac{WLCS(X,Y)}{f(m)} \right) \quad (12)$$

$f^{-1}$  adalah invers dari fungsi  $f$ .  $WLCS(X, Y)$  adalah bobot topik yang ditetapkan dan  $f(m)$  adalah jumlah seluruh kata (untuk bag of word) atau jumlah seluruh topik (untuk bag of phrase). Penghitungan tingkat akurasi topik yang berdasarkan bag of word menggunakan persamaan 13 sedangkan evaluasi untuk seluruh topik menggunakan persamaan 14.

$$R_{wlcs} = \frac{\sum_{w \in c_j} \log(1 + count_{match}(w))}{\sum_{w \in c_j} (count(w))} \quad (13)$$

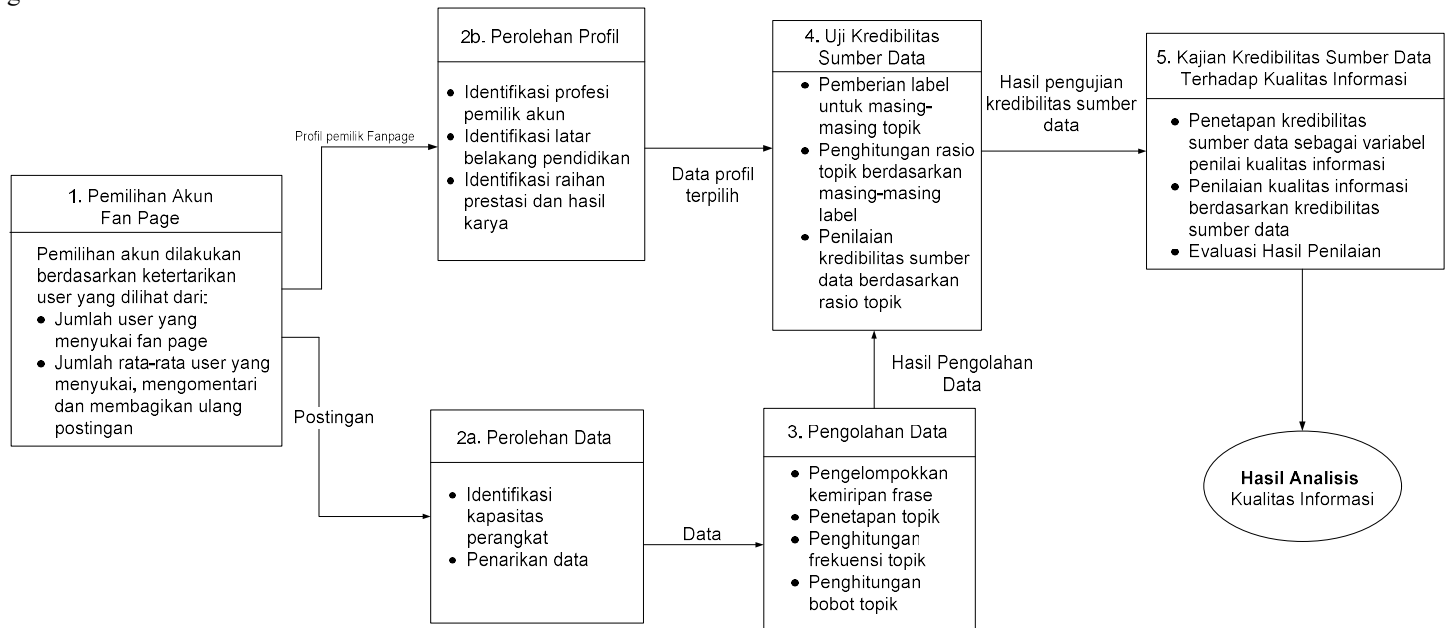
Keterangan:  $\sum_{w \in c_j} \log(1 + count_{match}(w))$  adalah total jumlah bobot setiap kata yang menyusun seluruh topik dan  $\sum_{w \in c_j} (count(w))$  adalah total kata pada seluruh data

$$R_{wlcs} = \frac{\sum_{w \in c_j} \log(1 + count_{match}(t))}{\sum_{w \in c_j} (count(t))} \quad (14)$$

Keterangan:  $\sum_{w \in c_j} \log(1 + count_{match}(t))$  adalah total jumlah bobot seluruh topik dan  $\sum_{w \in c_j} (count(t))$  adalah total jumlah topik pada seluruh data.

### III. HASIL DAN PEMBAHASAN

Sebagaimana teori yang dikaji pada bagian II dapat disimpulkan bahwa, mengukur kualitas informasi dilakukan dengan menilai kredibilitas sumber data, sehingga model yang dibuat memiliki tujuan khusus (*special purpose model*). Setiap teori dan konsep yang dikaji pada bagian II disusun secara utuh dengan metode pengolahan kata, sehingga membentuk skenario penilaian yang terdapat pada gambar 2.



Gambar 3 Rancangan Model Penilaian Kualitas Informasi

Sosial media yang dijadikan sebagai objek penelitian adalah Facebook, dengan cara memilih beberapa *Fan Page* yang dikelola pemiliknya, kemudian menilai kredibilitasnya. Facebook dipilih karena merupakan sosial media yang populer serta layanan *post content* di Facebook dapat memuat banyak kata untuk satu topik bahasan, sehingga pembahasan suatu topik dapat dilakukan secara luas. Adapun penjelasan setiap tahapan pada gambar 1 adalah sebagai berikut:

1. Pihak yang dipilih untuk memperoleh informasi darinya harus memiliki informasi yang banyak [18]. Dalam hal ini, pihak yang dipilih untuk dijadikan objek penelitian adalah pemilik Fan Page Facebook yang sangat aktif dan sering menyebarkan informasi yang menarik banyak perhatian pengguna Facebook. Hal tersebut dapat dilihat melalui jumlah penggemar yang dimiliki dan angka rata-rata *post content* yang disukai, dikomentari, dan dibagikan ulang oleh penggemarnya.
- 2a. Berdasarkan teori yang dikaji pada bagian II.E *Post content* diperoleh melalui suatu layanan Facebook yang dapat meng-ekstrak data *post content* sebanyak yang diperlukan. Tentunya hal ini juga dibatasi oleh kapasitas perangkat yang digunakan oleh peneliti, karena data yang sangat banyak akan memberikan banyak beban pada perangkat. Pembatasan juga dapat disesuaikan dengan ranah waktu postingan sesuai kebutuhan penelitian.
- 2b. Berdasarkan teori yang dikaji pada bagian II.E Profil diperoleh melalui halaman *Fan Page* pada bagian “about” meliputi latar belakang pendidikan, bidang kepakaran yang dimiliki, karya-karya yang pernah dihasilkan, dan penghargaan yang pernah diraih. Profil dapat diperoleh dari sumber informasi lain (jika diperlukan).

3. Berdasarkan teori yang dikaji pada bagian III penyimpulan topik konten dilakukan dengan cara mengolah seluruh kata dari *post content* yang diperoleh sebelumnya, sehingga menghasilkan informasi tentang topik yang sering dibahas. Jika merujuk kepada penelitian Fogg & Tseng (1999) yang membahas tentang *expertise*, maka dapat dipahami bahwa *post content* yang disebar pengguna sosial media mengindikasikan pengetahuan, pengalaman, dan kompetensi yang dimiliki olehnya. Langkah-langkah yang dilakukan dengan metode persiapan pengolahan kata, LDA, *Clustering*, Penyimpulan topik, dan evaluasi penyimpulan topik.
4. Tahapan ke-4 dilakukan dengan cara membandingkan topik yang sering dibahas dengan profil yang diperoleh. Perbandingan menghasilkan label kepada semua topik bahasan. Label *on-topic* diberikan kepada topik bahasan yang sesuai dengan keahlian, label *off-topic* diberikan kepada topik yang membahas kehidupan sehari-hari dan isu personal, label *cross-topic* diberikan kepada topik yang bahasannya diluar keahlian [16]. Penilaian kredibilitas dilakukan dengan memperhatikan rasio jumlah topik berdasarkan label yang diberikan.
5. Berdasarkan pernyataan Fogg & Tseng (1999) yang dituliskan pada bagian III dapat disimpulkan bahwa, kualitas informasi ditentukan berdasarkan kredibilitas sumber data. Jika sumber data dinilai kredibel, maka informasi yang dihasilkan dinilai berkualitas. Tetapi jika sumber data dinilai tidak kredibel, maka informasi yang disampaikan tidak berkualitas.

#### IV. PENUTUP

Model penilaian kualitas informasi pada penelitian ini merupakan *special purpose model*, yaitu model yang khusus untuk menilai kualitas informasi dari Fan Page Facebook. Tahapan pada gambar 2 dibentuk berdasarkan hasil kajian studi literatur dan analisis fenomena penggunaan Fan Page Facebook yang masih perlu dilakukan pembuktian terkait dijalankannya langkah-langkah yang ada pada model penilaian kualitas informasi di Fan Page Facebook.

#### REFERENSI

- [1] M. J. Kargar , “Towards Dedicated Information Quality Frameworks in Social Networks products”, International Conference on Uncertainty Reasoning and Knowledge Engineering, 2011.
- [2] S. Sikdar, et al., “Understanding Information Credibility on Twitter”, DOI 10.1109/SocialCom.2013.9., 2013.
- [3] D. Wang, “Analysis and Detection of Low Quality Information in Social Network”, IEEE, 2014.
- [4] R. Narayanam, et al., “Discovering Signature of Social Networks with Application to community Detection”, IEEE, 2014.
- [5] I. Lukmana, et. al., “Multi-Document Summarization Based on Sentence Clustering Improved Using Topic Words”, JUTI – Vol 12: 1 – 8, 2014.
- [6] D. M. Blei, et. al., “Latent Dirichlet Allocation”, Journal of Machine Learning Research 3: 993-1022, 2003.
- [7] I. Guy, et al., “Same places, same things, same people?: mining user similarity on social media”, Proceedings of the 2010 ACM conference on Computer supported cooperative work (pp. 41-50). Savannah: ACM, 2010.
- [8] Z. Papacharissi, “A Networked Self: Identity, Community, and Culture on Social Network Sites”, Routledge, New York, 2011.
- [9] S. Alter, 2002, “Information System : The Foundation of E-Business”, 4th edition, Prentice Hall, New Jersey, 2002.
- [10] L. Al-Hakim, “Challenges of Managing Information Quality in Service Organization”, Idea Group Inc., Hershey, 2006.
- [11] J. R. Talburt, “Entity Resolution and Information Quality”, Elsevier, Burlington, 2011.
- [12] Y. W. Lee, et. al., “AIMQ: a methodology for information quality assessment”, Information & Management, vol. 40, pp. 133 – 146, 2002.
- [13] B. Otto, “Towards a Framework for Corporate Data Quality Management”, 18th Australasian Conference on Information Systems, 2007.
- [14] F. Naumann, et. al., “Assessment methods for information quality criteria”, Proceedings of the 2000 Conference on Information Quality, 2000.
- [15] B.J. Fogg, “The Elements of Computer Credibility”, CHI '99 Proceedings of the SIGCHI conference on Human Factors in Computing Systems, Pages 80 – 87, New York, 1999.
- [16] K.R. Canini, et. al., “Finding Credible Information Sources in Social Networks Based on Content and Social Structure”, IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, 2011.
- [17] 01.ibm.com/support/knowledgecenter/SSEPGG\_8.2.0/com.ibm.datatools.datamining.doc/c\_similarity\_threshold.html; diakses tanggal 13 November 2015
- [18] Y. Shuang, “Effects of Information Quality and Source Credibility on EWOM Adoption in Context of Virtual Community”, International Conference on Management Science & Engineering, 2013.
- [19] D. Gujraniya, et al., “Efficient clasificaiton using phrases generated by topic model”, International Conference on Pattern Recognition, Tsukuba, Japan, 2012.
- [20] R.B. Yates dan B. R. Neto, “Modern Information Retrieval”, New York, Addison-Wesley, 1999.
- [21] EMC Education Services, “Data Science and Big Data Analytics: Discovering, Analyzing , Visualizing and Presenting Data”, John Wiley & Sons, Inc., Indianapolis, Indiana, 2015.
- [22] E. Setiawan.. KBBI Freeware. [Online]. Available: <http://pusatbahasa.diknas.go.id/kbbi/> , 2010.
- [23] I.P.G.H. Suputra et. al., “Pendekatan Positional Text Graph untuk Pemilihan Kalimat Representatif Cluster pada Peringkasan Multi-Dokumen”, Jurnal Ilmiah Ilmu Komputer, 2013.
- [24] C.Y. Lin, “ROUGE: A Package for Automatic Evaluation of Summaries”, In Proceedings of Workshop on Text Summarization Brances, Bancelona, hal. 74-81, 2004.