

Klasifikasi Keluhan Masyarakat Terhadap Layanan Publik pada Harian Radar Tarakan

Indra Tri Saputra, Program Studi Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati Tarakan.

Abstract— *Website* koran harian Radar Tarakan memiliki kolom dengan judul “Warga Menulis” di mana menu ini merupakan sarana bagi pembaca untuk menyampaikan keluhan ataupun aspirasi mereka. Yang menjadi permasalahan, pesan pembaca atau opini yang ditampilkan bersifat *to the point*, hanya isi opini sesuai yang dikirim pembaca tanpa informasi tambahan kepada siapa opini tersebut ditujukan. Tujuan dari penelitian ini adalah melakukan klasifikasi data opini pada *website* koran harian Radar Tarakan khususnya opini yang berkaitan dengan fasilitas dan pelayanan publik. Klasifikasi merupakan suatu proses pengelompokan data sesuai dengan kelas atau kategori yang telah ditentukan sebelumnya. Hipotesis yang dapat diambil adalah hasil klasifikasi diharapkan memiliki akurasi hingga 70%. Tahap awal dari proses klasifikasi yaitu *preprocessing* di mana pada tahap ini hal-hal yang dilakukan antara lain *case folding*, *tokenizing*, *convert word*, *stopword removal (filtering)* dan *stemming*. Algoritma yang digunakan dalam penelitian ini adalah Frequency Ratio Accumulation Method (FRAM). Pembuatan aplikasi menggunakan bahasa pemrograman PHP dan *database* MySQL. Hasil uji coba dari penelitian ini menunjukkan rata-rata akurasi yang diperoleh pada proses klasifikasi opini menggunakan algoritma FRAM adalah 60%. Besar kecilnya prosentase akurasi tergantung dari jumlah data latih yang digunakan. Semakin banyak jumlahnya dapat meningkatkan nilai akurasi akan tetapi hal ini akan berpengaruh terhadap efisiensi kinerja sistem.

Kata Kunci— FRAM, Klasifikasi, Opini, Machine Learning

I. PENDAHULUAN

Pemanfaatan teknologi informasi khususnya internet perlu dilakukan secara maksimal guna memenuhi kebutuhan masyarakat akan informasi yang up to date. Kebutuhan masyarakat akan informasi secara cepat dan mudah mendorong media-media penyedia informasi menyajikan informasi secara online. Salah satu yang memanfaatkan kelebihan internet tersebut adalah surat kabar harian atau koran.

Koran harian Radar Tarakan memiliki *website* dengan alamat www.kaltara.procal.co. Seperti *website* surat kabar pada umumnya, secara keseluruhan menu yang ada pada *website* www.kaltara.procal.co berisi informasi berita baik lokal, nasional, internasional maupun informasi penting lainnya.

Selain informasi berita terdapat satu kolom dengan judul “Warga Menulis” di mana menu ini merupakan sarana bagi pembaca untuk menyampaikan keluhan ataupun aspirasi mereka khususnya yang berhubungan dengan fasilitas dan pelayanan publik.

Pesan pembaca atau opini yang ditampilkan bersifat *to the point*, hanya isi opini sesuai yang dikirim pembaca tanpa informasi tambahan kepada siapa opini tersebut ditujukan. Sementara tidak semua opini secara terbuka menyertakan hal tersebut. Atau bisa saja pembaca keliru mengalamatkan opininya. Padahal opini berisi hal-hal yang ingin disampaikan pembaca mengenai permasalahan yang dihadapi dengan harapan akan ada penyelesaian atau tanggapan untuk permasalahan tersebut. Belum adanya penekanan kepada siapa opini tersebut ditujukan bisa berakibat opini tidak akan tersampaikan dengan benar. Untuk itu perlu adanya penekanan yang jelas agar keluhan pembaca tidak salah sasaran.

Sebuah disiplin ilmu *text mining* dapat digunakan untuk mengatasi permasalahan tersebut. *Text mining* [1] adalah suatu proses untuk mengambil informasi dari teks yang ada. *Text mining* mencari pola-pola yang ada di teks dalam bahasa natural yang tidak terstruktur. Algoritma yang digunakan dalam penelitian ini adalah Frequency Ratio Accumulation Method (FRAM). Algoritma FRAM merupakan salah satu metode pembelajaran *supervised document classification*. Metode klasifikasi FRAM adalah metode klasifikasi yang menggunakan jumlah rasio frekuensi pada tiap kategori dari fitur term individual.

Langkah yang akan diambil pada penelitian diawali dengan memasukkan opini dari alamat *website* koran harian Radar Tarakan www.kaltara.procal.co ke dalam *database*, kemudian dilakukan tahapan *preprocessing*. Langkah selanjutnya melakukan perhitungan menggunakan metode FRAM untuk memperoleh klasifikasi opini yang tepat sebelum ditampilkan pada alamat *website* www.kaltara.procal.co. Hasil akhir penelitian ini nantinya secara otomatis akan menampilkan opini berikut kelas atau kategorinya berdasarkan isi dari opini sesuai klasifikasi yang sudah tersedia.

II. TINJAUAN PUSTAKA

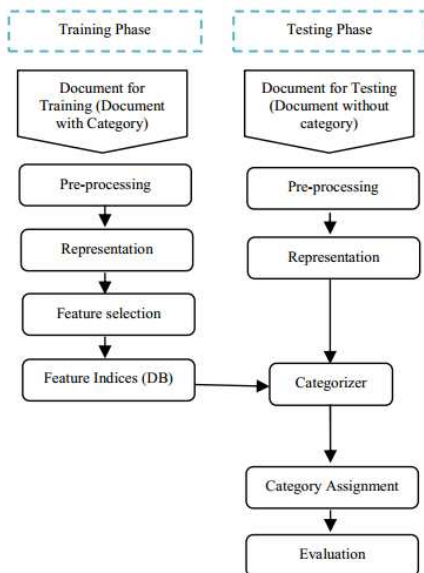
Kategorisasi teks [2] berkaitan dengan area informasi dan dokumentasi pengetahuan. Karena informasi dan pengetahuan disimpan dan dibagi ke dalam beberapa kategori atau teks, kategorisasi teks membantu pengguna dalam hal informasi dengan mengarahkan ke informasi yang diinginkan. Sebagian besar teknik kategorisasi teks seperti *Machine Learning* [3], [4],

Naïve Bayes (NB), K-Nearest Neighbour (KNN) [5], Support Vector Machine (SVM) dan Decision Tree memiliki model matematika yang kompleks [6]. Semuanya cenderung berkaitan dengan seleksi fitur dan categorization successively menjadi permasalahan tersendiri dalam kategorisasi teks otomatis, yang menyebabkan biaya dan masalah komputasi yang kompleks [7], [8].

Frequency Ratio Accumulation Method (FRAM) merupakan teknik klasifikasi sederhana yang menambahkan rasio dari frekuensi term pada masing-masing kategori, dan bisa menggunakan indeks term tanpa batas [9]–[12].

Metode ini ditandai sebagai klasifikasi dokumen tanpa mengekstraksi fitur term pada tahap seleksi fitur. Tugas kategorisasi dikombinasikan dengan tugas pengolahan fitur. Perbedaan tahap-tahap main building blocks antara kategorisasi teks otomatis standar dengan kategorisasi teks otomatis metode FRAM ditunjukkan pada Gambar 1 dan Gambar 2.

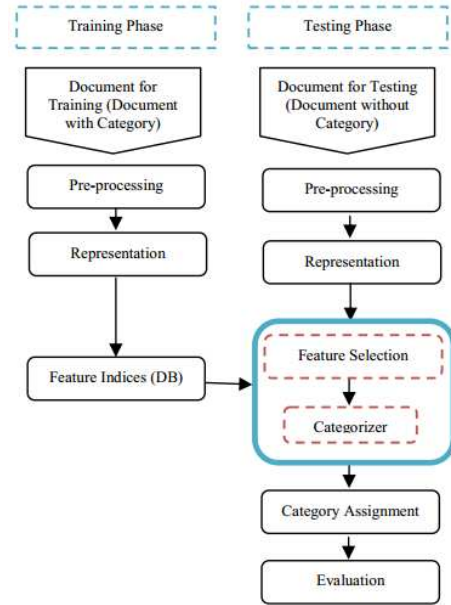
Tahapan proses baik itu pada kategorisasi teks otomatis standar maupun kategori teks otomatis FRAM terdiri dari 2 (dua) proses yaitu yang pertama adalah tahap pelatihan dimana dokumen yang diberi label di bawah kategori yang telah ditetapkan awalnya dilakukan inisialisasi pra-proses untuk menghilangkan term yang tidak berguna dan mengganggu. Selanjutnya, fitur term yang penting menjadi kata kunci diekstrak dalam tahap pelatihan mulai representasi hingga proses seleksi fitur dan menghasilkan database indeks, selanjutnya disebut Database (DB), yang kemudian digunakan untuk tahap yang kedua yaitu tahap uji. Pada tahap uji, pemeriksaan klasifikasi akan dievaluasi dengan mengelompokkan satu set dokumen pra-kategori satu demi satu sebagai dokumen tanpa kategori, dan kemudian mengukur kinerja kategorisasi menggunakan beberapa teknik standar evaluasi kinerja.



Gambar. 1. Kategorisasi Teks Otomatis Standar

Perbedaan antara kategorisasi teks otomatis FRAM pada Gambar 2 dengan kategori teks otomatis standar pada Gambar

1 terjadi pada tahap pertama yaitu tahap pelatihan dimana pada kategorisasi teks otomatis FRAM proses seleksi fitur akan dikecualikan dan proses seleksi fitur akan ditetapkan kategorinya pada tahap klasifikasi.



Gambar. 2. Kategorisasi Teks Otomatis FRAM

Hal ini yang menjadi kelebihan dari kategorisasi teks otomatis FRAM dimana dengan tidak adanya proses seleksi fitur pada tahap pelatihan akan mengurangi waktu komputasi yang berdampak pada kinerja sistem menjadi lebih efisien.

Metode ini pada awalnya menghitung total rasio frekuensi (FR) fitur term individual dalam tiap kategori sebagai berikut:

$$FR(t_n, c_k) = \frac{R(t_n, c_k)}{\sum_{c_k \in C} R(t_n, c_k)} \quad (1)$$

Di mana, rasio R tiap fitur term untuk masing-masing kategori dihitung dengan:

$$R(t_n, c_k) = \frac{f_{c_k}(t_n)}{\sum_{t_n \in T} f_{c_k}(t_n)} \quad (2)$$

$f_{c_k}(t_n)$:
Total Frequency fitur t_n di dalam kategori c_k

Dengan demikian, pada tahap pelatihan, rasio frekuensi dari semua fitur term dihitung dan dikumpulkan dalam tiap kategori. Selanjutnya, menghitung nilai evaluasi kategori atau skor kategori, di mana indikasi kemungkinan dokumen kandidat pada tahap uji termasuk dalam suatu kategori sebagai berikut:

$$E_{d_i}(c_k) = \sum_{t_n \in d_i} FR(t_n, c_k) \quad (3)$$

Akhirnya, dokumen kandidat d_i diklasifikasikan ke dalam kategori c_k di mana skor kategorinya maksimum, seperti berikut:

$$c_{\hat{k}} = \arg \max_{c_k \in C} E_{d_i}(c_k) \tag{4}$$

Metode yang diusulkan yaitu FRAM mempertahankan rasio frekuensi dalam tahap pelatihan dengan total jumlah fitur term yang di lambangkan dengan N dan total jumlah kategori yang di lambangkan dengan K. Selain itu, skor kategori untuk tiap kategori dihitung dengan menambahkan rasio frekuensi dari dokumen kandidat pada tahap uji mencakup fitur term dan golongan fitur term ke dalam kategori terkait yang mana skor evaluasinya maksimum.

III. METODE DAN INTI PENELITIAN

Algoritma dalam penelitian ini terdiri dari 2 (dua) tahap, yaitu tahap preprocessing data dan klasifikasi data opini.

A. Praproses Data

Pada tahap ini dilakukan serangkaian proses untuk mempersiapkan data. Seperti telah dijelaskan sebelumnya bahwa data opini yang digunakan masih berupa data teks tidak terstruktur (unstructured data), sehingga diperlukan langkah-langkah mempersiapkan data tersebut agar siap digunakan untuk proses selanjutnya.

1. Case Folding

Tahap pertama adalah case folding yang bertujuan mengubah teks ke bentuk huruf kecil semua serta menghilangkan karakter seperti tanda baca dan angka.

2. Tokenisasi

Tahap berikutnya tokenisasi yaitu memisahkan deretan kata pada kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal (termmed word).

3. Convert Word

Tahap selanjutnya convert word, yaitu merubah kata yang tidak baku menjadi lebih baku.

4. Stopword Removal (Filtering)

Tahap berikutnya adalah stopwords removal atau filtering, yaitu penghapusan term yang tidak berhubungan (irrelevant) dengan subyek utama dari database meskipun kata tersebut sering muncul dalam dokumen, misalnya: ada, adalah, agak, agar, akan dan sebagainya.

5. Stemming

Tahap terakhir adalah menghilangkan awalan dan akhiran kata atau stemming. Dengan stemming setiap kata dikembalikan ke bentuk dasarnya.

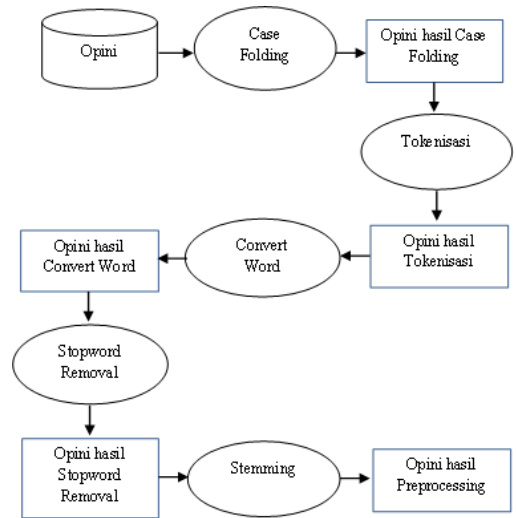
Adapun alur dari preprocessing data dapat dilihat pada Gambar 3.

B. Klasifikasi Data Opini

Klasifikasi menggunakan metode Frequency Ratio Accumulation Method (FRAM) melalui beberapa tahapan, yaitu:

1. Pembagian Data Latih dan Uji

Setelah tahap praproses data (preprocessing), dataset dibagi menjadi 2 (dua) bagian, data latih (training) dan data uji (testing). Data latih digunakan sebagai data pembelajaran untuk menemukan pola yang tepat terhadap sekumpulan dataset, sehingga saat proses testing diperoleh hasil klasifikasi yang tepat. Pembagian data dilakukan dengan cara mengambil berapa persen dataset untuk training dan berapa persen untuk testing, misalnya 70% data training dan 30% data testing. Pembagian data dilakukan secara acak dari keseluruhan dataset yang digunakan.



Gambar. 3. Preprocessing Data

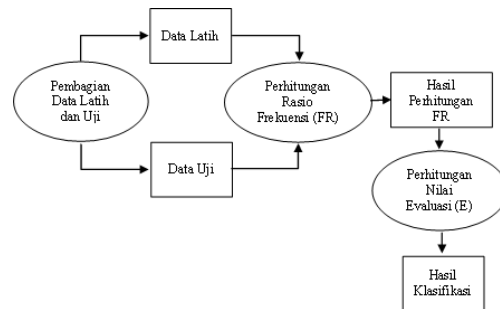
2. Perhitungan Rasio Frekuensi (FR)

Kemudian data latih akan dibuat dalam bentuk matriks, di mana baris mewakili kata atau term dan kolom mewakili dokumen yang dikelompokkan berdasarkan kelas atau kategorinya masing-masing. Nilai matriks $D(n,k)$ adalah jumlah kemunculan kata n pada kelas k , seperti ditunjukkan pada persamaan (2)

3. Perhitungan Nilai Evaluasi (E)

Tahap akhir yaitu mencari nilai evaluasi (E) menggunakan persamaan (3)

Adapun alur dari klasifikasi data opini menggunakan metode FRAM dapat dilihat pada Gambar 4.



Gambar. 4. Klasifikasi Opini Metode FRAM

C. Metode Evaluasi

Cross Validation merupakan salah satu teknik untuk menilai atau memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Dalam teknik ini dataset dibagi menjadi sejumlah K-buah partisi secara acak. Kemudian dilakukan sejumlah K-kali eksperimen, di mana masing-masing eksperimen menggunakan data partisi ke-K sebagai data testing dan memanfaatkan sisa partisi lainnya sebagai data training. Sebagai gambaran, jika kita melakukan 5-Fold Cross Validation maka desain data eksperimennya sebagai berikut.

Dataset: K1, K2, K3, K4, K5

TABLE I.
DATA EKSPERIMEN

<i>Eksperimen Ke</i>	<i>Data Testing</i>	<i>Data Latih</i>
1	K1	K2,K3,K4,K5
2	K2	K1,K3,K4,K5
3	K3	K1,K2,K4,K5
4	K4	K1,K2,K3,K5
5	K5	K1,K2,K3,K4

IV. HASIL EKSPERIMEN DAN PENELITIAN

Algoritma dalam penelitian ini diimplementasikan menggunakan bahasa pemrograman PHP dengan database MySQL. Prosesnya diawali dengan memasukkan opini dari alamat website koran harian Radar Tarakan www.kaltara.procal.co ke dalam database, kemudian dilakukan tahapan preprocessing. Langkah selanjutnya penentuan jenis kelas untuk data opini baru dengan menggunakan metode klasifikasi FRAM.

Perbandingan antara proses klasifikasi dengan cara manual ataupun menggunakan metode FRAM dengan jumlah dataset 500 opini belum mencapai hasil sebagaimana tertulis dalam hipotesis penelitian. Perbandingan antara proses klasifikasi manual dengan klasifikasi metode FRAM menggunakan teknik validasi *Cross Validation* rata-rata tingkat akurasi yang diperoleh adalah 51%.

Ketidaksesuaian antara kelas aktual dengan kelas prediksi pada proses klasifikasi lebih disebabkan pada tidak berimbangnya jumlah data yang ada pada setiap kelas atau kategori. Semakin banyak jumlahnya dapat meningkatkan nilai akurasi akan tetapi hal ini akan berpengaruh terhadap efisiensi kinerja sistem. Daftar beberapa hasil klasifikasi opini baru dapat dilihat pada tabel II

TABLE II.
KESESUAIAN JENIS KELAS UNTUK DATA OPINI BARU

<i>No.</i>	<i>Opini</i>	<i>Kelas</i>	<i>Kesesuaian Jenis Kelas</i>
1	KEPADA DKPP, kenapa pembersihan jalan dilakukan pada siang hari, mengganggu pengguna jalan. Debunya di mana-mana	DKPP	Sesuai
2	DINAS perhubungan kabupaten Bulungan dan kota Tarakan, mohon diperhatikan perlengkapan keselamatan speedboat bagi penumpangnya. Bila perlu diperiksa terlebih dahulu perlengkapan-perengkapan tersebut, demi keselamatan	DISHUB	Sesuai
3	TARAKAN makin kacau, daging asal Malaysia dilarang masuk tapi daging Indonesia harganya selangit. Minyak tanah susah ditiap RT tapi pengecer di sepanjang jalan selalu banyak dapat minyak tanah & dijual dengan harga 45 ribu tiap 5 liter. Apa pemerintah Tarakan tidak bisa mengendalikan semua ini untuk kesejahteraan masyarakat, khususnya Tarakan?!	DISPERI NDAGK OP	Sesuai
4	Lagi Dinas PU Kabupaten Nunukan memperhatikan jalan dari Long Bawan sampai lagi ke Midang Kecamatan Krayan. Karena sangat rusak parah.	PEMDA	Berbeda
5	KEPADA perusahaan yang sudah diberi surat edaran tentang UMK, segera naikan gaji karyawan. Jangan sampai memberi gaji karyawan tidak sesuai ketentuan	DISOSN AKER	Sesuai

V. KESIMPULAN

Dari percobaan yang dilakukan pada proses klasifikasi data opini masyarakat menggunakan metode FRAM, hasil yang diperoleh untuk akurasi terbesar atau tertinggi adalah 60%. Perbandingan antara proses klasifikasi manual dengan klasifikasi metode FRAM menggunakan teknik validasi *Cross Validation* rata-rata tingkat akurasi yang diperoleh adalah 51%.

Ketidaksesuaian antara kelas aktual dengan kelas prediksi pada proses klasifikasi lebih disebabkan pada tidak berimbangnya jumlah data yang ada pada setiap kelas atau kategori. Perbandingannya adalah jumlah data terbesar ada pada kelas atau kategori "PLN" dengan 124 data dan jumlah data terkecil ada pada kelas atau kategori "PENGADILAN" dengan 9 data. Sedangkan rata-rata jumlah data pada setiap kelas adalah 45 data.

DAFTAR PUSTAKA

- [1] R. Feldman, J. Sanger, and others, *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press, 2007.
- [2] R. J. Mooney, "Machine Learning Text Categorization," *Austin Univ. Texas*, 2006.
- [3] Fitria, Gunawan, and E. I. Setiawan, "Abstract Summarization Using Maximum Marginal Relevance and Vector Space Model."
- [4] M. Suzuki, N. Yamagishi, T. Ishida, M. Goto, and S. Hirasawa, "On a new model for automatic text categorization based on vector space model," in *2010 IEEE International Conference on Systems, Man and Cybernetics*, 2010, pp. 3152–3159.
- [5] A. Indriani, E. Novianto, and others, "Weight Adjusted K-Nearest Neighbor dan Minimum Spanning Tree untuk Information Retrieval System di Perpustakaan STMIK PPKIA Tarakanita Rahmawati Tarakan," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 2013, vol. 1, no. 1.
- [6] G. Pant and P. Srinivasan, "Learning to crawl: Comparing classification schemes," *ACM Trans. Inf. Syst.*, vol. 23, no. 4, pp. 430–462, 2005.
- [7] J. Wang and A. An, "Classification Methods," 2005, pp. 144–149.
- [8] A. Mahinovs, A. Tiwari, R. Roy, and D. Baxter, *Text classification method review*. 2007.
- [9] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," 2006.
- [10] M. Suzuki and S. Hirasawa, "Text categorization based on the ratio of word frequency in each categories," in *2007 IEEE International Conference on Systems, Man and Cybernetics*, 2007, pp. 3535–3540.
- [11] M. Suzuki, T. Ishida, and M. Goto, "Refinement of index term set and improvement of classification accuracy on text categorization," in *2008 International Symposium on Information Theory and Its Applications*, 2008, pp. 1–6.
- [12] B. T. Sharef, N. Omar, and Z. T. Sharef, "An automated arabic text categorization based on the frequency ratio accumulation.," *Int. Arab J. Inf. Technol.*, vol. 11, no. 2, pp. 213–221, 2014.