

FRAMEWORK SISTEM REKOMENDASI BERITA BERBAHASA INDONESIA BERDASARKAN PILIHAN MINAT BACA PERSONAL

Diana Purwitasari, Rizki Akbar Maulana, Ary Mazharuddin Shiddiqi

Jurusan Teknik Informatika, Fakultas Teknologi Informasi

Institut Teknologi Sepuluh Nopember

Kampus ITS Sukolilo Surabaya 60111

Telp. +6231 5939214, Fax. +6231 5913804

E-mail: diana@if.its.ac.id

Abstract

Indonesian news recommendation system is implemented using the data in a personal preferences of news to meet the needs of news readers in Indonesia, which has been increasing year to year. The implementation of this system we get different news recommendation for every reader. The algorithm, that is used, is hybrid filtering as a combination of collaborative filtering and content-based filtering. Collaborative filtering is an algorithm that is often used in recommendation system. This algorithm recommends item or news based on the items selected by other users. Content-based filtering is a recommendation algorithm that provides content of items that most preferred by the user. The highest recommended value is given to items that have the highest value in a collaborative filtering recommendation and content-based filtering. An experiment that is performed is compare two simulations that run news recommendation in recommending the news with a number of different categories. The test result showed 44% of news recommended category were news reader favorite.

Abstrak

Menjamurnya penyedia situs layanan berita Indonesia menciptakan kebutuhan untuk dapat memberikan rekomendasi berita sesuai dengan ketertarikan masing-masing individu. Rekomendasi pada pembaca dengan pendekatan Collaborative Filtering diberikan berdasarkan berita-berita yang biasa diakses oleh sekelompok orang dengan minat baca sama. Sedangkan rekomendasi dengan pendekatan Content-based Filtering akan menganalisis kategori favorit pembaca berdasarkan pilihan berita yang disukai di masa lalu. Rekomendasi Collaborative Filtering dilakukan menggunakan weblog pada server penyedia situs layanan yang mencatat kunjungan pembaca pada berita-berita tersedia. Kemudian kelompok pembaca berdasarkan kunjungannya atau minat baca personalnya diidentifikasi dengan algoritma pengklasteran hierarkikal Agglomerative Complete-Linkage. Pada Content-based Filtering akan memperhitungkan ketertarikan pembaca dan kategori berita yang sedang populer. Gabungan dari kedua metode rekomendasi tersebut atau diistilahkan dalam makalah ini sebagai Hybrid Filtering mengacu pada algoritma perkomendasi berita yang digunakan oleh Google News. Namun rekomendasi berita dengan Google News belum dapat diberikan untuk berita berbahasa Indonesia. Makalah ini berisi bahasan tentang framework sistem rekomendasi berita dengan pendekatan Hybrid Filtering untuk merekomendasikan berita berbahasa Indonesia dengan cara menganalisis pilihan pembaca secara personal. Percobaan dilakukan dengan mempersiapkan purwarupa situs layanan berita berbahasa Indonesia yang memiliki fitur perkomendasi. Berdasarkan hasil percobaan dari log seminggu kunjungan pengguna situs penyedia berita terlihat bahwa keberagaman ketertarikan pembaca mempengaruhi kemungkinan ditemukannya berita yang benar-benar disukai untuk direkomendasikan.

Kata kunci: rekomendasi berita, collaborative filtering, content-based filtering

1. PENDAHULUAN

Kebiasaan membaca berita di Indonesia berubah dengan perkembangan dalam teknologi Internet. Menjamurnya penyedia situs layanan berita Indonesia menciptakan kebutuhan untuk dapat memberikan rekomendasi berita sesuai dengan ketertarikan masing-masing individu. Salah satu cara memberikan rekomendasi berita dapat

dilakukan secara personal dengan menyarankan berita tersebut kepada teman dekat melalui Yahoo! Messenger (YM) atau Yahoo! News. Selain itu Google Inc. telah meluncurkan media untuk merekomendasikan berita disebut Google News yang menggunakan data berita dari layanan berita situs berbahasa Inggris (Das, dkk., 2007). Namun rekomendasi berita dengan Google News belum dapat diberikan untuk

berita berbahasa Indonesia. Google News saat itu (2007) menggunakan pendekatan pemilihan berita (*news filtering*) disebut *Collaborative-based Filtering* yaitu rekomendasi berdasarkan berita yang biasa diakses oleh sekelompok orang dengan minat baca sama. Metode *news filtering* tersebut dianggap masih kurang mewakili keinginan minat dari pembaca sehingga ditambahkan pendekatan yang lain dalam memberikan rekomendasi berita yaitu *Content-based* dan *Collaborative Filtering* (Liu, dkk., 2010). Rekomendasi dengan pendekatan *Content-based Filtering* akan menganalisis kategori favorit pembaca berdasarkan pilihan berita yang disukai di masa lalu. Makalah ini membahas gabungan dari kedua metode rekomendasi milik Google News atau diistilahkan disini sebagai *Hybrid Filtering* untuk merekomendasikan berita berbahasa Indonesia yang menjadi kontribusi utamanya. Proses *Collaborative-based Filtering* dilakukan menggunakan weblog pada server penyedia situs layanan yang mencatat kunjungan pembaca pada berita – berita yang disediakan. Meskipun makalah ini juga mengambil konsep *Collaborative-based Filtering* namun hal yang berbeda adalah adanya proses identifikasi pengelompokkan orang dengan minat baca sama menggunakan algoritma pengklasteran hierarkikal yaitu *Agglomerative Complete-Linkage* (Manning, dkk., 2008).

2. METODOLOGI

Pengerjaan penelitian ini dimulai dari pemrosesan berita, lalu penerapan algoritma filtering untuk rekomendasi berita, setelah itu dilanjutkan dengan proses uji coba.

2.1 Pemrosesan Berita

Sistem Rekomendasi Berita dapat dianggap sebagai layanan satu pintu untuk mendapatkan berita-berita dari berbagai situs penyedia berita *online*. Berita sebagai data utama dalam sistem rekomendasi diunduh berdasarkan tautan pada file RSS (*Really Simple Syndication*) yang sebelumnya telah ditentukan (Rajaraman & Ullman, 2011). Sistem memiliki modul unduh berita yang secara otomatis mengambil berita – berita terbaru. Namun hierarki pengelompokkan berita pada situs tidak selalu sama. Oleh karena itu berita yang diunduh diberikan label kategori secara otomatis oleh Sistem Rekomendasi Berita.

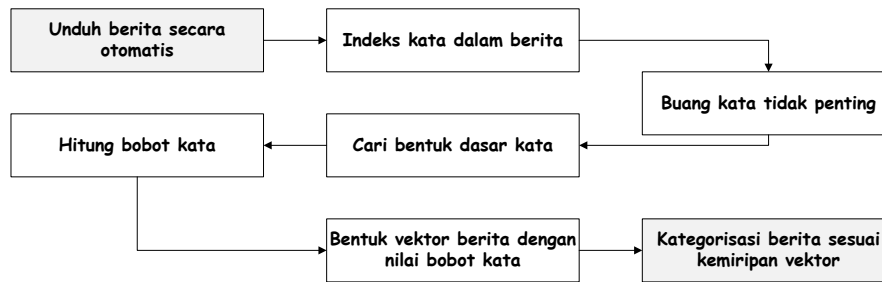
Pemberian label secara otomatis membutuhkan analisa teks dari isi berita. Tahapan pemrosesan teks dalam berita ditunjukkan pada Gambar 1 yaitu antara lain adanya proses *stemming*, penghitungan bobot sampai kemiripan isi (Montes-y-Gómez, dkk., 2001).

Stemming merupakan pengolahan kata untuk mendapatkan kata dasar dari sebuah kata yang telah mengalami imbuhan dengan asumsi bahwa kata-kata tersebut memiliki makna sama pula. Pada dasarnya jenis imbuhan kata dalam Bahasa Indonesia terbagi tiga yakni imbuhan awalan (*prefiks*), akhiran (*sufiks*), dan sisipan (*infiks*). Salah satu pengembangan algoritma *stemming* berbahasa Indonesia (*ECS -- Enhanced Confix Stripping Stemmer*) adalah pencarian kata dasar dari kata berimbuhan (Arifin, dkk., 2009). Pendekatan *Confix Stripping Stemmer* akan memastikan langkah awal adalah penghilangan akhiran atau awalan yang memberikan hasil kata dasar lebih baik. Proses pemenggalan kata-kata dengan bentuk perulangan juga dimungkinkan dengan keberadaan kamus kata dasar yang lebih lengkap. Kemudian ECS melengkapi aturan pemenggalan awalan dibanding algoritma *stemmer* Bahasa Indonesia yang sudah ada khususnya kesalahan dalam pemenggalan akhiran.

Untuk mengurangi jumlah kata yang akan mengalami *stemming* (cari bentuk dasar kata) maka kata – kata sambung dan semacamnya seperti “dan, yang, ke, jika” dibuang terlebih dahulu. Kata-kata tidak penting sering disebut sebagai *stopwords*. Kemudian jumlah kata yang menjadi isi berita atau terdaftar atau terindeks menjadi semakin berkurang lebih lanjut akibat proses *stemming*. Selanjutnya setiap kata pasti memiliki bobot atau tingkat kepentingan kata. Kata yang sering muncul pada satu berita lebih penting dibanding kata-kata yang jarang. Nilai kemunculan kata disebut sebagai *term frequency* (TF). Tetapi kata yang muncul di setiap berita juga memiliki sedikit nilai signifikansi. Oleh karena itu kemunculan kata di setiap berita, disebut *indeks document frequency* (IDF), juga tidak boleh terlalu tinggi. Nilai bobot kata dihitung dari hasil perkalian TF dan IDF untuk kata tersebut sesuai dengan Persamaan (1) (Manning, dkk., 2008).

$$\omega_{ij} = tf_{ij} \times idf_i \quad (1)$$

Nilai ω_{ij} adalah bobot kata ke- i (k_i) pada dokumen berita ke- j (d_j). Sedangkan nilai tf_{ij} menunjukkan banyaknya kata ke- i yang muncul dalam dokumen berita ke- j . Selanjutnya nilai idf_i yaitu banyaknya berita yang memiliki kata ke- i sebagai isinya. Hasil dari pemrosesan berita adalah representasi vektor berita M_j dengan nilai bobot kata menjadi isi selnya, $m_{ij} = \omega_{ij}$. Sel m_{ij} menunjukkan bobot kata ke- i pada dokumen berita ke- j . Vektor representasi berita akan digunakan dalam pengelompokkan berita berdasarkan isinya.



Gambar 1. Tahapan dalam Pemrosesan Berita.

Untuk pemberian label secara otomatis maka sebelumnya dilakukan kegiatan semacam pembelajaran oleh admin Sistem Rekomendasi dengan mengidentifikasi secara manual setiap berita sesuai kategori yang telah ditentukan. Kemudian sistem akan mencari “wakil berita” yang menjadi ciri khas untuk kategori tersebut dengan menghitung nilai tengah (*center*) dari vektor berita dalam kategori menggunakan Persamaan (2).

$$v_{ic} = \left[\frac{\sum_{d_j \in c} m_{ij}}{\#d_j \in c} \right] \quad (2)$$

Vektor “wakil berita” memiliki dimensi atau banyak kata sama dengan dimensi pada vektor berita. Nilai sel v_{ic} atau bobot kata ke-*i* untuk “wakil berita” dari kategori ke-*c* adalah nilai rata – rata bobot kata ke-*i* dari semua dokumen yang ada dalam label kategori ke-*c*. Dokumen berita ke-*j* dikelompokkan ke dalam kategori ke-*c* berdasarkan nilai kemiripan tertinggi dari isi dokumen berita (d_j) dengan isi “wakil berita” (v_c). Nilai kemiripan tersebut dihitung menggunakan model *cosine similarity* pada Persamaan (3) yaitu kedekatan antar dua vektor yang diukur dari nilai *cosine* sudut pembentuknya (Manning, dkk., 2008).

$$kemiripan(d_j, v_c) = \frac{d_j \cdot v_c}{\|d_j\| \|v_c\|} \quad (3)$$

2.2 Algoritma Filtering untuk Rekomendasi Berita

Rekomendasi berita merupakan hasil dari pemilihan berita (*filtering*) berdasarkan aturan tertentu: *collaborative-based*, *content-based*, atau gabungan keduanya.

Collaborative-based Filtering

Collaborative-based Filtering memperhatikan rating dari pengguna-pengguna yang memiliki sifat tidak jauh berbeda dengan pengguna yang hendak diobservasi. Oleh karena itu pengguna harus dikelompokkan berdasarkan kesamaan dalam membaca berita-berita serupa. Tingkat kesamaan antar pengguna, misal usr_A dan usr_B , dihitung berdasarkan *Jaccard Coefficient* (De Meo, dkk., 2011) sesuai dengan Persamaan (4).

$$kesamaan(usr_A, usr_B) = \frac{\# \text{ berita sama dibaca kedua pengguna}}{\# \text{ berita dibaca } usr_A + \# \text{ berita dibaca } usr_B} \quad (4)$$

Pengguna dengan minat baca personal sama dikelompokkan dengan algoritma pengklasteran hierarkikal *Agglomerative Complete Linkage* (Manning, dkk., 2008) berikut ini:

- buat matriks kunjungan pengguna dengan baris menyatakan pengguna dan kolom menyatakan berita yang dikunjungi
- inisialisasi jumlah kluster awal pengguna berdasarkan minat sebanyak pengguna yang tercatat
- buat matriks jarak antar kluster yang isinya dihitung menggunakan Persamaan (4)
- gabung dua kluster yang memiliki jarak terdekat dan buat matriks jarak antar kluster yang baru dengan anggota kluster adalah hasil gabungan
- ulangi langkah c dan d sampai jumlah kluster terbentuk sesuai dengan jumlah yang diinginkan

Pemilihan pengklasteran hierarkikal disebabkan minat baca memungkinkan untuk terbagi secara berjenjang. Semisal pembaca yang tertarik dengan politik dalam negeri namun tidak pada politik luar negeri tetap dapat direkomendasikan berita politik. Dua kluster digabung menjadi satu apabila penggabungan memberikan tingkat kesamaan paling minimal untuk setiap pasang pengguna antar kluster. Nilai kesamaan antar pengguna dinyatakan pada Persamaan (4). Pengelompokkan pengguna berdasarkan minat baca personal atau identifikasi profil dilakukan secara periodikal untuk memperbaharui data.

Langkah-langkah pemberian rekomendasi berita untuk pengguna usr_A pada waktu ke *t* menggunakan *Collaborative Filtering* (Liu, dkk., 2010) adalah: diasumsikan bahwa profil pembaca yang diperbaharui telah tersedia setiap waktu tertentu (misalnya pukul 00.00 WIB saat rendahnya kemungkinan adanya kunjungan dari pembaca).

Profil akan diidentifikasi dengan algoritma hierarkikal *Agglomerative Complete-Linkage*.

Profil pengguna usr_A dikenali/dikelompokkan kedalam profil yang sudah ada berdasarkan nilai kesamaan Jaccard di Persamaan (4) dengan melihat log akses berita yang telah tercatat dari t -hari. Jumlah hari pada uji coba ditentukan sebanyak 6 hari. Setelah ditemukan profil pembaca tersebut maka diambil n berita yang

$$skor_{con}(d_j) = \frac{p(\text{berita di kategori } d_j)_{1 \text{ jam}} \times \left(\frac{\#pembaca d_j \times p(\text{pembaca berita di kategori } d_j)}{p(\text{berita di kategori } d_j)_{6 \text{ hari}}} + \#klik \text{ virtual} \right)}{\#pembaca d_j + \#klik \text{ virtual}} \quad (6)$$

Content-based Filtering

Pada hakikatnya pembaca memiliki ketertarikan yang tetap. Seorang pembaca yang memiliki ketertarikan terhadap berita kategori politik, untuk selanjutnya cenderung tetap tertarik terhadap berita tersebut. Namun tidak dapat dipungkiri tren berita populer juga menarik perhatian pembaca. Sebagai contoh pada saat Olimpiade Beijing berlangsung jumlah peminat berita olahraga meningkat tajam.

Langkah-langkah pemberian rekomendasi berita untuk pengguna usr_A pada suatu waktu tertentu t menggunakan *Content-based Filtering* (Liu, dkk., 2010) adalah sebagai berikut: rekomendasi dokumen berita ke- j dengan *Content-based Filtering* akan memiliki $skor_{con}(d_j)$ yang dihitung sesuai dengan Persamaan (6). Semisal dengan kategori berita d_j adalah kategori ke- c , maka nilai $p(\text{berita di kategori } d_j)_{1 \text{ jam}}$ yaitu banyaknya berita yang muncul dalam kategori ke- c pada $t =$ satu jam terakhir terhitung dari saat pengguna meminta rekomendasi berita. Nilai $p(\text{berita di kategori } d_j)_{6 \text{ hari}}$ adalah nilai peluang banyaknya berita yang muncul dalam kategori ke- c pada $t =$ enam hari terakhir.

Selanjutnya nilai $\#pembaca d_j$ menunjukkan banyaknya akses dari pengguna yang membaca dokumen berita d_j . Kemudian $\#klik \text{ virtual}$ berfungsi sebagai jaring pengaman seandainya dokumen tersebut belum pernah dibaca oleh pengguna (pada uji coba diset =10). Lalu nilai $p(\text{pembaca berita di kategori } d_j)$ merupakan peluang banyaknya pengguna yang membaca berita-berita di kategori ke- c yaitu kategori dari berita d_j sesuai dengan log akses sistem.

2.3 Hybrid Filtering

Untuk rekomendasi berita bagi pengguna usr_A dengan *Hybrid Filtering* (Liu, dkk., 2010) maka dokumen berita ke- j yang direkomendasikan akan memiliki nilai $skor_{hyb}(d_j)$ dihitung dengan Persamaan (7).

sering dikunjungi oleh pengguna dalam profil. Pada uji coba nilai n ditentukan sebanyak 9 berita. Rekomendasi berita diberikan sesuai urutan $skor_{col}(d_j)$ yaitu tingkat keseringan dokumen berita ke- j dibaca pada Persamaan (5) (Liu, dkk., 2010).

$$skor_{col}(d_j) = \#kunjungan \text{ baca berita } d_j \quad (5)$$

$$skor_{hyb}(d_j) = skor_{col}(d_j) \times skor_{con}(d_j) \quad (7)$$

Penghitungan nilai $skor_{hyb}(d_j)$ pada Persamaan (7) terdiri dari tiga tahap (Liu, dkk., 2010). Tahap pertama pilih set berisi n_1 dokumen berita dengan *Collaborative-based Filtering* disebut Set1. Pada tahapan ini profil pengguna usr_A juga diketahui. Berita pada Set1 sudah memiliki $skor_{col}(d_j)$. Kemudian hitung $skor_{con}(d_j)$ untuk berita pada Set1. Tahap Kedua pilih set berisi n_2 dokumen berita dengan *Content-based Filtering* disebut Set2. Berita pada Set2 sudah memiliki $skor_{con}(d_j)$. Kemudian hitung $skor_{col}(d_j)$ untuk berita pada Set2. Nilai $skor_{col}(d_j)$ pada Set2 bisa bernilai nol jikalau dokumen berita d_j tidak pernah dikunjungi oleh pembaca dalam profil pengguna usr_A . Tahap ketiga gabungkan Set1 dan Set2 dengan banyak dokumen maksimal adalah $n_1 + n_2$ berita. Untuk setiap berita hitung $skor_{hyb}(d_j)$ sesuai dengan Persamaan (7).

2.4 Uji Coba

Pengujian sistem rekomendasi berita dilakukan pada dua buah komputer yang berperan sebagai *server* dan *client*. Komputer *server* untuk Sistem Rekomendasi Berita dan komputer *client* untuk pengguna. Setiap komputer memiliki prosessor Intel(R) Core(TM)2 Duo CPU E7400 @ 2.80GHz, memori 2038MB serta browser Mozilla Firefox 3.6 untuk *client*.

Skenario pengujian adalah sebagai berikut:

- buat satu situs berita dengan berita yang sudah diberi label topik sebagai kelasnya untuk pelatihan
- buat representasi vector untuk setiap berita
- hitung nilai vector wakil kelas suatu topik berita
- beri label topik pada berita baru dengan hitung kedekatan vector berita baru dan vector wakil kelas berita menggunakan Persamaan (3)
- catat kunjungan pengguna yang membaca berita di situs tersebut

Jakarta - PT Jasa Marga Tbk (JSMR) tengah mengembangkan teknologi transaksi terbaru di pintu tol yang membuat pengguna jalan bahkan tidak perlu membuka jendela untuk melakukan transaksi.

Menurut Direktur Utama Jasa Marga Frans Sunito, dengan adanya teknologi tersebut, maka transaksi di pintu tol tidak akan memakan waktu lama sehingga antrian pun tidak memanjang.

... ..

Selain itu, perusahaan pelat merah itu juga sedang mengupayakan agar para pengguna E-Toll Card tidak hanya bisa menggunakannya untuk membeli bensin dan berbelanja di Indomaret dan Solaria saja, tetapi juga bisa untuk transaksi parkir.

"Kita upayakan kerjasama dengan operator parkir. Kalau nanti sudah bisa ini akan sangat menolong, tidak hanya untuk beli bakmie di Solaria saja," jelasnya.

Gambar 2. Cuplikan Isi Berita.

- f. kluster minat pembaca berdasarkan log kunjungan berita dengan Persamaan (4)
- g. buat representasi vector untuk setiap kluster minat pembaca teridentifikasi
- h. hitung skor berita yang memperhitungkan informasi minat pembaca dengan Persamaan (6)
- i. buat rekomendasi baca untuk
 - (1) pengunjung baru di situs
 - hitung kedekatan vector berita yang sedang dibaca dan setiap vector wakil kelas berita teridentifikasi dengan Persamaan (3)
 - pilih satu kelas berita berdasarkan jarak kedekatan terpendek antar vector berita dan vector wakil kelas berita
 - tampilkan berita-berita dalam kelas tersebut diurut berdasarkan skor berita
 - (2) pengunjung aktif di situs
 - buat representasi vector pengunjung berdasarkan log berita yang dibaca
 - hitung kedekatan vector pengunjung dan setiap vector wakil kluster minat pembaca yang teridentifikasi dengan Persamaan (3) juga
 - pilih satu kluster minat pembaca dengan jarak kedekatan terpendek
 - tampilkan berita-berita yang dibaca oleh pengunjung dalam kluster tersebut diurut berdasarkan skor berita

Implementasi pengujian dilakukan dengan cara simulasi pembacaan berita selama enam hari dan kemudian pada hari ke tujuh akan direkomendasikan berita. Situs berita memiliki sepuluh berita baru per hari sehingga hasil rekomendasi dilakukan di hari ketujuh akan menganalisa sekitar 60 berita tersimpan. Data-data berita diambil dari situs Detik.com selama dua minggu.

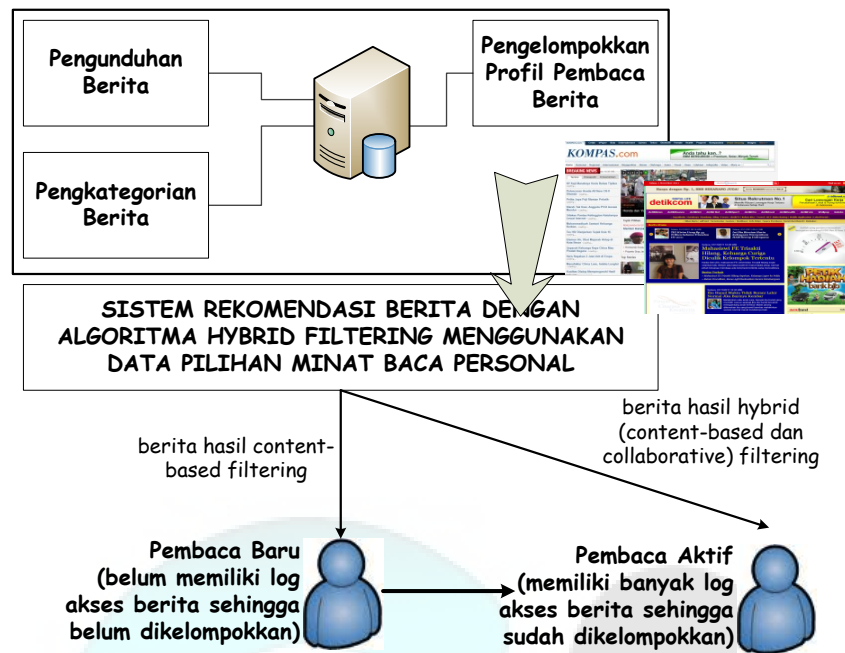
Asumsi pertama yang digunakan adalah seorang pembaca hanya memiliki satu profil saja semisal profil pembaca ibu muda (usia 25-30 tahun) yang menyukai berita gosip seputar artis. Asumsi kedua adalah sebuah berita hanya

memiliki satu label kategori. Misalkan berita pada Gambar 2 dikategorikan berita Bisnis namun sebenarnya bisa menjadi kategori Hiburan karena berita tersebut membahas adanya teknologi baru dan terdapat kata-kata berkonteks pemasaran seperti transaksi, belanja, antrian, atau parkir.

Pengguna dari sistem rekomendasi berita berbahasa Indonesia dibedakan berdasarkan minat bacanya yaitu pengguna baru, pengguna pasif dan pengguna aktif yang ditunjukkan pada Gambar 3. *Pengguna Baru* adalah pengguna yang baru terdaftar pada sistem dan ingin menggunakan fitur rekomendasi. Sistem rekomendasi tersebut terkait erat dengan ketersediaan berita online dari situs penyedia berita. Kemudian untuk pengguna yang banyak membaca berita sehingga memiliki catatan log akses berita disebut sebagai *Pengguna Aktif*.

Situs untuk pengujian memiliki berita terbagi menjadi tiga kategori: olahraga, bisnis, hiburan. Kunjungan 10 pembaca dicatat untuk proses rekomendasi yang melakukan pembacaan sebanyak kurang lebih 4 – 7 berita perharinya. Akses berita lebih banyak dilakukan pada berita-berita di kategori olahraga dan bisnis. Uji coba menunjukkan bahwa rekomendasi berita untuk kategori olahraga dan bisnis menghasilkan berita yang sesuai sekitar 65%.

Artinya 6 dari 9 rekomendasi berita yang dilakukan di hari ketujuh sesuai dengan kategorinya. Namun rekomendasi berita untuk kategori hiburan memiliki akurasi 0% karena kunjungan berita yang tercatat di enam hari sebelumnya sebagian besar adalah berita pada kategori olahraga dan bisnis. Namun berita yang direkomendasikan dapat dikatakan memiliki nilai “hiburan” tersendiri karena berita tersebut secara kontekstual dapat dikelompokkan sebagai berita hiburan; misal berita berjudul “Teknologi Baru, Bayar Tol Tanpa Buka Jendela Mobil”.



Gambar 3. Diagram Kontekstual untuk Implementasi Sistem Rekomendasi Berita.

5. SIMPULAN dan SARAN

Makalah ini berisi bahasan tentang algoritma *Hybrid Filtering* (gabungan metode *Content-based* dan *Collaborative Filtering*) untuk merekomendasikan berita berbahasa Indonesia dengan cara menganalisis pilihan pembaca secara personal. Percobaan dilakukan dengan mempersiapkan purwarupa situs layanan berita berbahasa Indonesia yang memiliki fitur perkomendasi. Hasil menunjukkan bahwa kemungkinan ditemukan berita yang benar – benar disukai dan memang berita tersebut direkomendasikan dipengaruhi keberagaman ketertarikan pembaca. Sehingga untuk menghasilkan rekomendasi berita yang lebih baik diperlukan data pembaca atau kelompok pembaca yang mencukupi.

6. DAFTAR RUJUKAN

Arifin, A.Z., Mahendra, I.P.A.K, Ciptaningtyas, H.C., 2009. Enhanced Confix Stripping Stemmer and Ants Algorithm for Classifying News Document in Indonesian Language. In: *the 5th Intl. Conf. on Information & Communication Technology and Systems (ICTS 2009)*. Surabaya, Indonesia

- Das, A., Datar, M., Garg, A., Rajaram, S., 2007. Google News Personalization: Scalable Online Collaborative Filtering. In: *16th Intl. Conf. on World Wide Web (WWW 2007)*. Banff, Alberta, Canada 8-12 May 2007. New York, NY, USA.
- De Meo, P., Ferrara, E., Fiumara, G., 2011, *Social Networking and Community Behavior Modeling: Qualitative and Quantitative Measurement*, Chapter: Finding Similar Users in Facebook, Igi Publishing
- Liu, J., Dolan, P., Pedersen, E.R., 2010. Personalized News Recommendation Based on Click Behavior. In: *the 15th Intl. Conf. on Intelligent User Interfaces (IUI 2012)*. Hong Kong, China
- Manning, C.D., Raghavan, P., Schütze, H., 2008. *Introduction to Information Retrieval*. Cambridge University Press
- Montes-y-Gómez, M., Gelbukh, A., López-López, A., 2001. Mining the News: Trends, Associations, and Deviations. *Computacion y Sistemas*, 5(1), pp.14-24
- Rajaraman, A., Ullman, J.D., 2011. *Mining of Massive Datasets*, Cambridge University Press