

**PERANCANGAN SISTEM REKOMENDASI ARTIKEL
ONLINE DENGAN *CONTENT BASED FILTERING*
(Brilio.net)**

SKRIPSI

Digunakan Sebagai Syarat Maju Ujian Diploma IV
Politeknik Negeri Malang

Oleh:

MUHAMMAD IQBAL ROFIKURRAHMAN. NIM. 1941727019



**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
JULI 2020**

BAB III. METODELOGI PENELITIAN

3.1 Data

Data yang akan diolah merupakan data yang berasal dari artikel yang telah dibuat sebelumnya pada situs brilio.net (data uji), pengolahan data berfokus kepada isi artikel seperti judul, gambar, isi konten, tanggal terbit dan alamat URL. Dalam proses perhitungan algoritma, semua data kecuali alamat URL, gambar, dan tanggal terbit akan digabung menjadi satu dokumen untuk dibandingkan dengan dokumen yang telah dibuat sebelumnya.

Akan ada proses pengambilan data secara terjadwal pada sistem untuk memperoleh data baru pada situs brilio.net, data baru yang diperoleh akan di catat jumlahnya dalam sebuah *history*.

Tabel 3. 1 Contoh data hasil crawl

Thumbnail	Title	Date	Url
https://cdn-brilio-net.akamaized.net/news/2020/05/13/184420/gara-gara-corona-mira-lesmana-ungkap-produser-film-rugi-miliaran-200513t.jpg	Mira Lesmana sebut produser film sudah rugi miliaran karena corona	2020-05-13	https://brilio.net/selebritis/gara-gara-corona-mira-lesmana-ungkap-produser-film-rugi-miliaran-200513t.html
https://cdn-brilio-net.akamaized.net/news/2020/05/13/184399/10-momen-kebersamaan-ranbir-kapoor-alia-bhatt-makin-mesra-200513v.jpg	10 Momen kebersamaan Ranbir Kapoor & Alia Bhatt, makin mesra	2020-05-13	https://brilio.net/selebritis/10-momen-kebersamaan-ranbir-kapoor-alia-bhatt-makin-mesra-200513v.html

Pada tabel 3.1 terdapat beberapa kolom yang tidak ditampilkan, sebagai contoh adalah kolom *content*, pada kolom content semua *text* akan diambil.

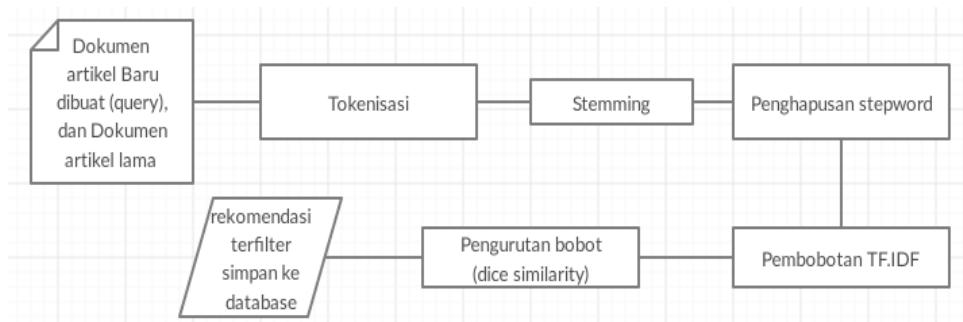
3.2 Metode Pengambilan Data

Metode yang digunakan dalam proses pengambilan data adalah metode *crawling*, dimana metode ini juga dikenal dengan metode *scrapping* yang intinya melakukan ekstraksi sebuah halaman pada situs tertentu untuk memperoleh data yang dibutuhkan, khususnya pada penelitian kali ini adalah situs *brilio.net*.

Metode ini memiliki tingkat kesulitan tersendiri tergantung situs yang akan dilakukan ekstraksi. Semua data yang diperoleh adalah data legal, dan semua data yang tampil saat membuka situs secara manual adalah data yang dapat dilakukan proses *crawling*.

3.3 Metode Pengolahan Data

Metode pengolahan data dimulai saat pertama kali data diperoleh dari proses *crawling*. Data yang akan diperoleh untuk diolah adalah judul artikel dan isi konten, yang telah diformat menjadi satu dokumen tanpa adanya tag html menjadi teks keseluruhan. Sebelumnya akan disediakan beberapa data sebagai data training untuk mengenalkan sistem.



Gambar 3. 1 Alur pengolahan data

Disaat ada artikel baru yang terbaca dari sistem *crawling* maka data tersebut akan dilakukan pemrosesan seperti proses *preprocessing* dan dilanjutkan dengan melakukan perhitungan pembobotan dengan TF.IDF, setelah menentukan bobot dari TF.IDF selanjutnya bobot tersebut akan dihitung nilai kedekatannya dengan metode *dice similarity* dimana metode ini adalah metode yang sering digunakan untuk mencari nilai kedekatan suatu dokumen.

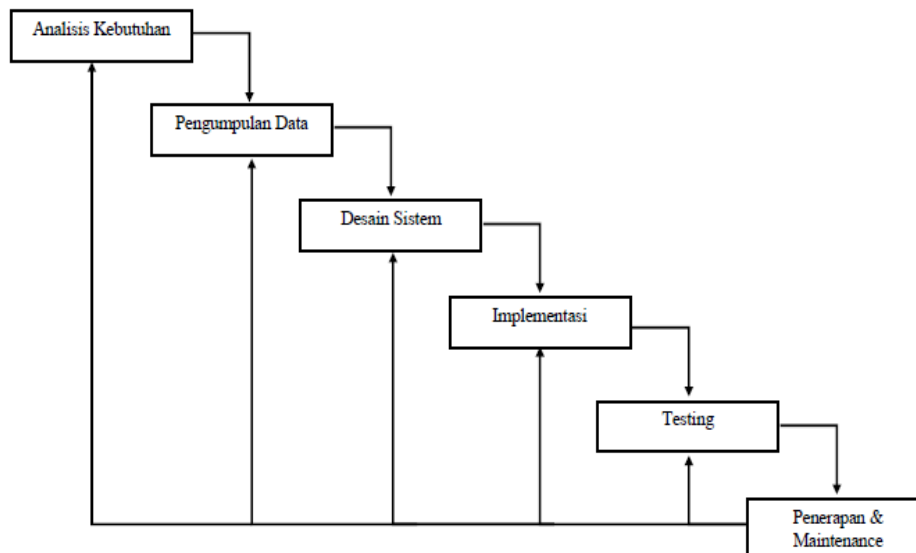
Keunggulan dari *dice similarity* dibandingkan metode lain yaitu dapat menentukan prioritas *precision* dan *recall* sesuai kebutuhan. Maka *dice similarity* menjadi metode yang dirasa tepat untuk menjalankan sebuah sistem rekomendasi artikel berita yang cepat.

Setelah data rekomendasi ditentukan dan difilter sesuai kebutuhan, selanjutnya data akan disimpan kedalam database dengan *id* sesuai dengan *id* artikel yang baru, dimana artikel baru ini digunakan sebagai *query* pembandingan dengan artikel lainnya.

Untuk mengatasi keterlambatan pembaharuan data rekomendasi, maka dibuatlah sebuah alur sistem dimana ketika artikel lama di akses kembali oleh pengguna, perhitungan rekomendasi akan dilakukan pembaharuan berdasarkan data artikel terbaru berdasarkan tanggal pembuatan artikel.

3.4 Metode Pengembangan Perangkat Lunak

Metode dalam penelitian ini menggunakan metode waterfall. Dalam metode ini terdapat 6 tahap, langkah-langkahnya adalah sebagai berikut:



Gambar 3. 2 Diagram Alur *Waterfall*

a. *Requirement Analysis*

Pada tahap ini akan dijelaskan tentang semua kebutuhan sistem untuk menjalankan sistem rekomendasi artikel ini, baik kebutuhan fungsional dan kebutuhan non fungsional.

b. Pengumpulan Data

Proses pengumpulan data dilakukan setiap hari dan dilakukan oleh sistem, sehingga data yang diolah adalah data dinamis yang sifatnya tidak tetap.

c. *Design*

Desain sistem akan terbagi menjadi beberapa bagian, seperti bagian admin atau sistem kontrol dan bagian rekomendasi atau desain untuk pengguna pada situs brilio.net

d. *Implementation*

Dalam tahap ini dilakukan pembuatan aplikasi. Pembuatan perangkat lunak dipecah menjadi modul-modul kecil yang nantinya akan digabungkan dalam tahap berikutnya. Selain itu dalam tahap ini juga dilakukan pemeriksaan terhadap modul yang dibuat, apakah sudah memenuhi fungsi yang diinginkan.

e. *Testing*

Pada tahap testing dilakukan uji coba pada aplikasi yang telah dibuat untuk mengetahui apakah perangkat lunak yang dibuat telah sesuai dengan desainnya dan apakah masih terdapat kesalahan

f. *Maintenance*

Tahap ini dilakukan jika perangkat lunak sudah jadi dan mengalami kerusakan. Maka akan dilakukan pemeliharaan. Pemeliharaan termasuk dalam memperbaiki kesalahan yang tidak ditemukan pada langkah sebelumnya.

3.5 Metode Pengujian Data

Pengujian yang dilakukan untuk menguji sistem dilakukan dengan 2 langkah, yaitu pengujian fungsionalitas dan pengujian performa sistem. Pengujian bertujuan untuk memastikan bahwa sistem yang dibangun telah berjalan sesuai dengan yang diharapkan.

a. **Pengujian Fungsionalitas**

Uji coba fungsionalitas dikatakan berhasil apabila fungsi yang ada pada sistem sesuai dengan yang diharapkan. Uji fungsionalitas yang diterapkan pada penelitian ini ialah dengan mencoba menjalankan setiap fitur yang ada dan yang disediakan pada sistem ini

b. **Pengujian Akurasi Sistem**

Dalam pengujian ini, dilakukan untuk mengetahui keakuratan aplikasi dalam menampilkan data hasil pencarian oleh pengguna

