Jurnal Transistor Elektro dan Informatika (TRANSISTOR EI)
Vol. 2, No. 1, Mei 2017 , pp. 31~39

Penerapan Algoritma *K-Nearest Neighbor* (KNN) dalam Pembuatan Sistem Penentuan Topik Artikel Berbasis Web

Nur Fatahna, Suryani Alifah, Sam Farisa C.
Teknik Informatika, Universitas Islam Sultan Agung

Correspondence Author: nurfatahna@std.unissula.ac.id

Abstrak
Penelitian ini bertujuan untuk menerapkan algoritma K-Nearest Neighbor dalam pembuatan sistem penentuan topik artikel.Algoritma K-Nearest Neighbor merupakan salah satu metode berbasis NN yang paling tua dan populer di dalam melakukan pengkategorian teks. Dalam penentuan prediksi label kelas pada data uji ditentukan dengan nilai k yang menyatakan jumlah tetangga terdekat. Dari k tetangga terdekat yang terpilih dilakukan voting dengan memilih kelas yang jumlahnya paling banyak sebagai label kelas hasil prediksi pada data uji. Klasifikasi dianggap sebagai metode terbaik dalam preses ketika data latih yang berjarak paling dekat dengan objek. Cara kerja dari KNN perlu adanya penetuan inputan berupa data latih, data uji dan nilai k.

Keyword: *Artikel, K-Nearest Neighbor, topik.*

1. PENDAHULUAN

Teknologi sangat bermanfaat untuk menunjang aktivitas yang dilakukan sehari-hari. Berkembangnya teknologi menjadikan manusia semakin mudah dalam berkomunikasi dan mendapatkan informasi. Salah satu informasi yang mudah didapatkan yaitu informasi yang dimuat dalam bentuk artikel.

Artikel merupakan suatu fakta yang dianalisis sehingga menimbulkan pendapat atau pandangan penulis atas fakta yang dianalisis atau bisa disebut juga opini yang disampaikan oleh seorang penulis tentang masalah aktual yang menyita perhatian masyarakat. Artikel berisi gagasan yang bertujuan memberitahu, mempengaruhi, meyakinkan, dan menghibur [1].

Penelitian terdahulu telah ada yang membahas mengenai aplikasi penentuan topik artikel menggunakan algoritma *K-Nearest Neighbor* (KNN). Salah satunya yaitu jurnal yang dibuat oleh Yoseph Samuel, dkk (2015) dalam jurnalnya dijelaskan bahwa dokumen berita akan dibuat pertopik agar pencarian berita sesuai topik yang diinginkan oleh pembaca lebih mudah. Algoritma *K-Nearest Neighbor* (KNN) biasanya menggunakan *majority vote* sebagai sebuah landasan penetuan dimana sebuah dokumen diklasifikasi. Namun disini, peneliti melakukan penggantian penggunaan *majority vote* menjadi *decision rule* dengan harapan agar penggunaan algoritma *K-Nearest Neighor* dapat dimaksimalkan. Pengambilan berita yang akan digunakan sebagai data training diambil dari tiga website yaitu :bbc.com, cnn.com, dan foxnews.com. berita-berita yang masuk tersebut akan dikategorikan berdasarkan topik olahraga yang terbagi menjadi 7 subtopik yakni : Soccer, Formula 1, Basketball, Motorsport, Baseball, Tennis, dan NFL. Sedangkan, berita yang digunakan sebagai data training sejumlah 280 berita dan untuk pengujian akan digunakan 95 berita baru [2].

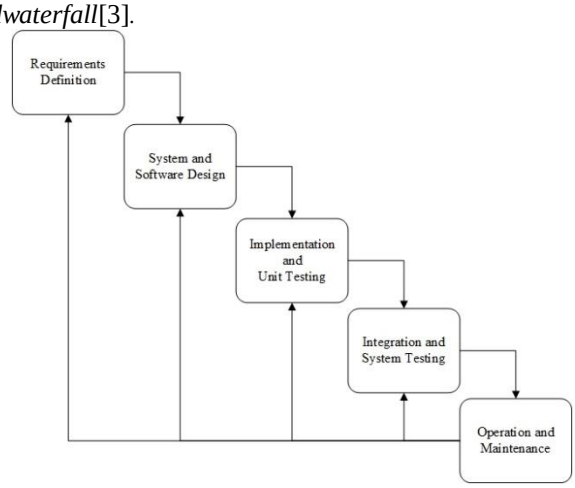
Berdasarkan beberapa penelitian terdahulu diatas, penulis akan membuat penelitian dengan menerapkan algortima *K-Nearest Negihbor* untuk menentukan topik sebuah artikel berdasarkan rumpun ilmu.

2. METODE PENELITIAN

Metodologi penelitian pada laporan tugas akhir ini adalah sebagai berikut :

a. Metode Pengumpulan Data
   Data-data diperlukan untuk mendukung pemecahan masalah yang timbul berdasarkan fokus penelitian. Data-data tersebut diperoleh dengan menggunakan metode studi literatur. Studi literatur merupakan cara yang dipakai untuk menghimpun data-data atau sumber-sumber yang berhubungan dengan topik yang diangkat dalam penelitian ini. Studi literatur bisa didapatkan dari berbagai sumber, seperti jurnal, buku dokumentasi, internet dan pustaka.
b. Model Proses Pengembangan Sistem

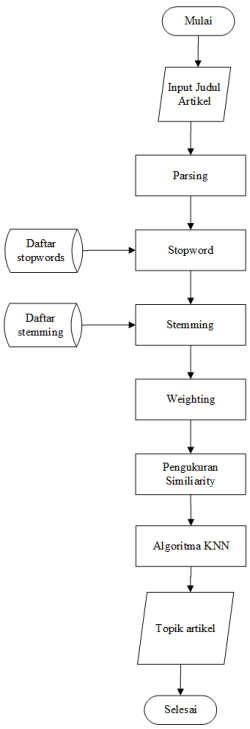
Proses dalam pengembangan sistem ini menggunakan model *modifiedwaterfall*. Berikut merupakan gambar model *modifiedwaterfall*[3].



Gambar 1 Model Proses Pengembangan Sistem

Berdasarkan model *waterfall* diatas, maka tahapan prosedur penelitian yang akan dilakukan adalah*Requirement Definition*, *System and Software Design*, *Implementation and Unit Testing*, *Integration and System Testing*, dan tahap*Operation and Maintenance*.

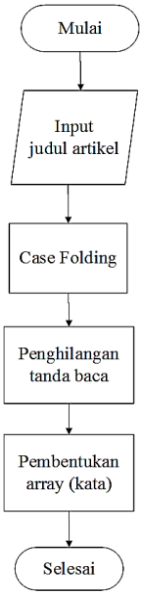
Sistem yang akan dibangun adalah sistem penentuan topik artikel yang menggunakan algoritma *K-Nearest Neighbor* . adapun *flowchart* sistem yang dibuat secara umumdapat dilihat pada gambar 2.



Gambar 2*Flowchart*sistem penentuan topik artikel

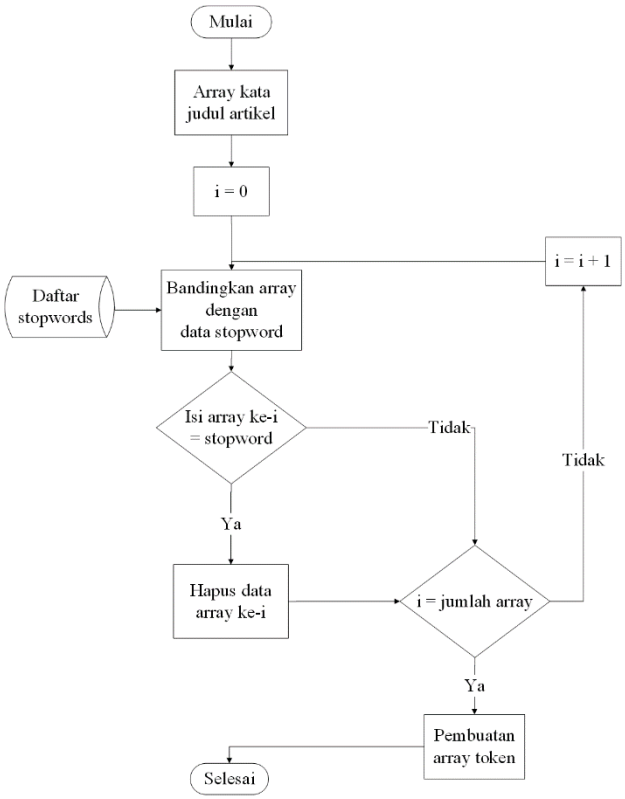
Gambar 2 menunjukkan proses keseluruhan sistem mulai dari penginputan judul artikel, proses parsing, penghilangn stopword, proses stemming, proses pembobotan (*weighting*), pengukuran similarity menggunakn *dicesimilarity*, algortitma KNN dan *output*-nya berupa topik artikel.

Tahapan proses sistem yang dijabarkan pada gambar 2 terdiri dari proses *pre-processing* yaitu parsing, *stopword*, dan *stemming*.



Gambar 3*Flowchart*proses parsing

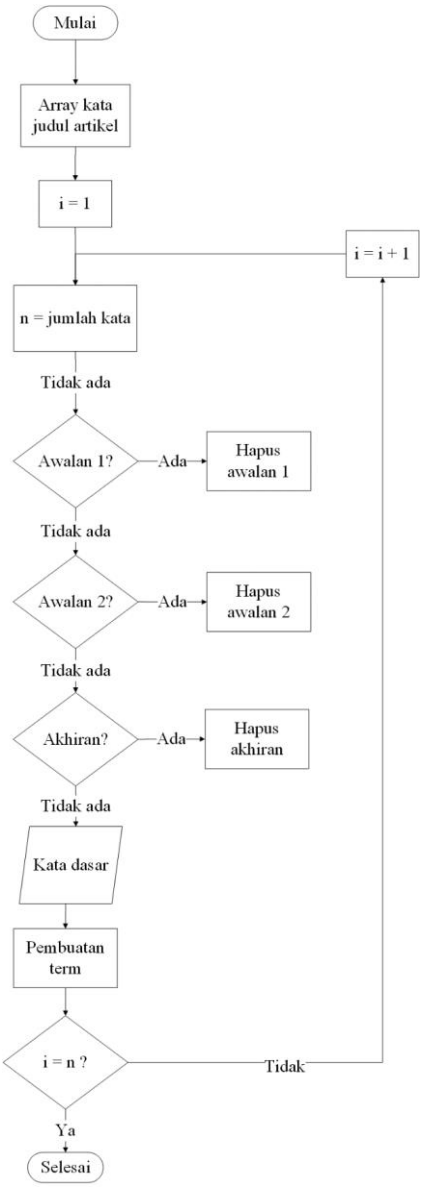
Gambar 3 merupakan *flowchart* proses *parsing*. Pada tahap awal proses *parsing* yaitu *case folding* yakni membuat semua huruf pada kalimat atau teks yang diinputkan menjadi huruf kecil, hal tersebut dilakukan untuk memperkecil ukuran *database* pada indeks. Kemudian proses selanjutnya yaitu penghilangan tanda baca seperti tanda koma (,), titik (.), dan lain sebagainya. Setelah itu teks yang diinputkan berupa kalimat dipisah menjadi bentuk kata atau *array*. *Output* yang dihasilkan pada proses ini yaitu sebuah *array* yang akan digunakan untuk proses berikutnya.



Gambar 4*Flowchartstopword*

Gambar 4 merupakan proses penghilangan *stopwords* yang dilakukan untuk menghilangkan kata yang dianggap tidak penting pada kalimat yang diinputkan. Baca *array* kata yang didapatkan dari proses *parsing*.

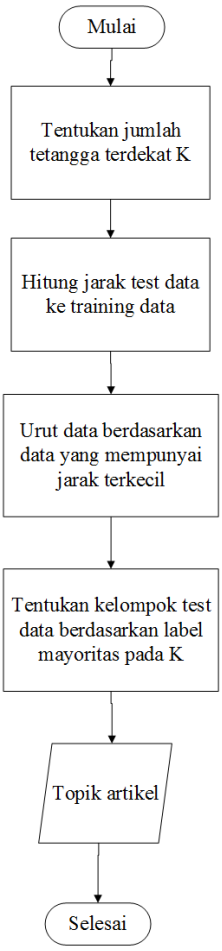
1. Ambil daftar *stopwords* pada *database*.
2. Bandingkan *array* dengan daftar *stopwords*. Jika kata termasuk ke dalam daftar *stopwords*, maka buang kata dari *array*.
3. Ulangi ke langkah ke tiga sampai *array* paling terakhir.



Gambar 5*Flowchartstemming*

Gambar 5 merupakan proses *stemming* yang dilakukan untuk menemukan kata dasar dari sebuah kata. Proses *stemming* dilakukan dengan menghilangkan semua imbuhan (afiks) baik yang terdiri dari awalan (prefiks) maupun akhiran (sufiks) . *Stemming* diidapatkan dari *array* kata hasil dari proses penghilangan *stopwords*. Algoritmanya sebagai berikut:

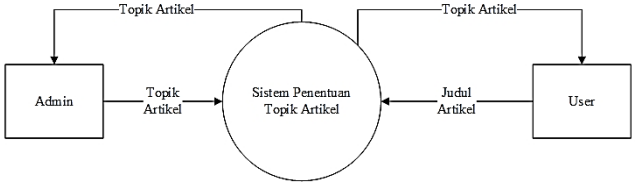
1. Baca *array* kata yang didapatkan dari proses penghilangan *stopwords*.
2. Hapus partikel dan hapus kata ganti
3. Apabila *array* memiliki awalan 1 maka hapus awalan 1, jika tidak ada, maka hapus awalan 2 lalu hapus akhiran.
4. Apabila *array* memiliki akhiran, maka hapus akhiran lalu hapus awalan 2
5. Selanjutnya mencocokkan dengan kamus kata dasar yang ada didalam *database*.
6. Ulangi langkah ke 2 sampai *array* kata paling terakhir.



Gambar 6 *Flowchart* KNN

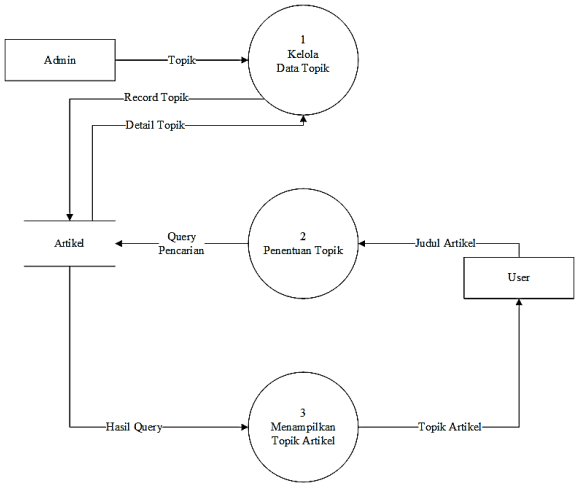
3. DIAGRAM KONTEKS

Diagram konteks menggambarkan gambaran umum proses yang ada pada sistem yang ditunjukkan pada gambar 7.



Gambar 7 Diagram Konteks

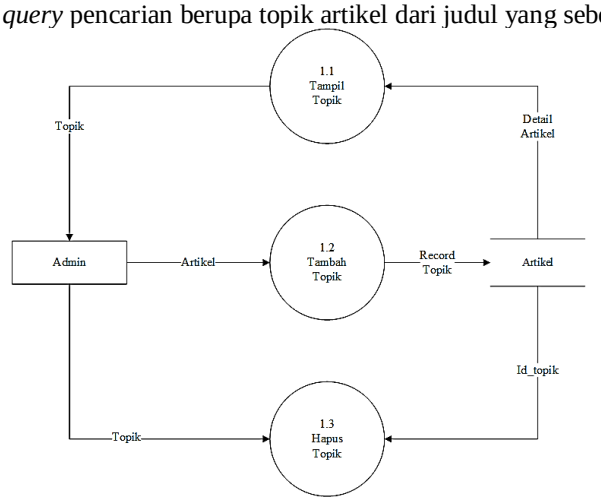
Sistem penentuan topik artikel melibatkan dua aktor yaitu admin dan *user*. Admin bertugas untuk mengelola sistem dan *user* bertindak sebagai pemakai sistem.



Gambar 8 DFD Level 0

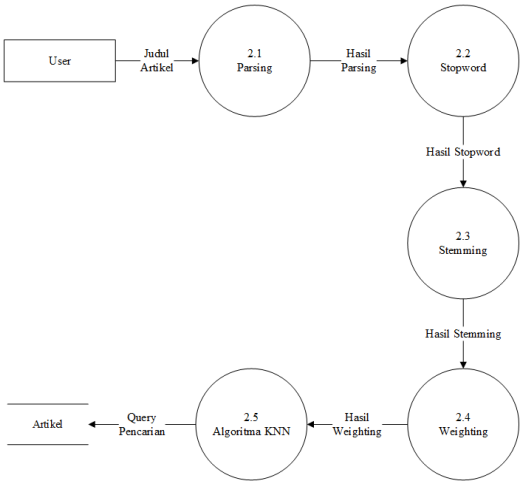
Gambar 8 menunjukkan aliran data sistem penentuan artikel yang melibatkan 2 aktor dan 3 proses. Berikut ini penjelasan dari masing-masing proses pada sistem :

1. Proses kelola topik di kelola langsung oleh admin. Admin bertugas untuk mengelola topik pada sistem yang kemudian disimpan pada tabel artikel.
2. Proses penentuan topik melibatkan *user*, dimana *user* berperan penting untuk menginputkan judul artikel kemudian akan diproses pada data proses dan hasil *query* pencarian dimasukkan pada tabel artikel.
3. Proses ketiga yaitu menampilkan topik artikel, dimana hasil *query* pencarian yang ada pada tabel artikel akan ditampilkan ke *user*. Hasil *query* pencarian berupa topik artikel dari judul yang sebelumnya diinputkan.



Gambar 9 DFD Level 1 proses pengelolaan topik

Pada gambar diatas proses pengelolaan topik dibagi menjadi 3 subproses yaitu tampil topik, tambah topik, dan hapus topik. Untuk subproses tampil topik aliran data dimulai dari mengambil data detail topik dari tabel artikel masuk ke proses tampil kemudian akan ditampilkan detail topik ke admin. Untuk subproses tambah topik aliran data dimulai dari admin memasukkan data topik kemudian di proses dan disimpan pada tabel artikel. Subroses selanjutnya yaitu hapus topik aliran data dimulai dari admin memilih topik yang akan dihapus kemudian dipanggil data tersebut dari tabel artikel kemudian dilakukan proses hapus artikel.

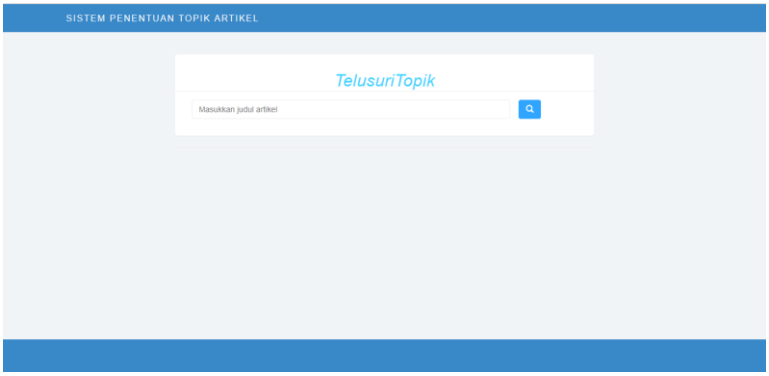


Gambar 10 DFD Level 1 Proses penentuan topik

Gambar 10 merupakan gambar aliran proses penentuan topik. Ada beberapa proses yang dilalui sebelum mendapatkan topik atau kategori yang dari judul artikel yang diinputkan yaitu *parsing*/tokenisasi, penghilangan *stopword*, *stemming*, pembobotan (*weighting*), perhitungan *similiarity*, dan yang terakhir perhitungan menggunakan algoritma KNN. Dimulai dari *user* menginputkan judul artikel, selanjutnya melalui tahap *prepocessing* dimana judul yang diinputkan akan melalui proses *parsing* atau tokenisasi yaitu proses pemotongan *string* input berdasarkan tiap kata penyusunnya. Setelah dilakukan tokenisasi selanjutnya digunakan fungsi *stopword*. Fungsi *stopword* untuk menghilangkan kata-kata yang dianggap tidak diperlukan. Selanjutnya kalimat dipisah menjadi kata perkata dan dilakukan proses *stemming* untuk mencari kata dasar dari kalimat tersebut. Setiap dokumen yang masuk akan dihitung bobotnya dengan cara menghitung berapa kali kata tersebut muncul dalam sebuah kalimat, kemudian akan dicocokkan dengan *query* yang diinputkan. Perhitungan tersebut menggunakan rumus tf (*termfrequency*) dan idf (*index document frequency*). Langkah selanjutnya setelah dilakukan proses *parsing* hingga pembobotan yaitu mengukur kemiripan kata yang akan diuji digunakan rumus *dice similiarity*. Setelah kata dasar ditemukan, berdasarkan kata dasar tersebut akan dikelompokkan sesuai kategori masing-masing dari tiap kata menggunakan *clustering* dan algoritma *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* bekerja berdasarkan jarak terpendek dari *sample* uji ke *sample* latih untuk menentukan nilai KNN. Dari kata yang telah masuk dalam kategori dipilih kata yang paling banyak dengan kategori yang sama yang menjadi topik artikel.

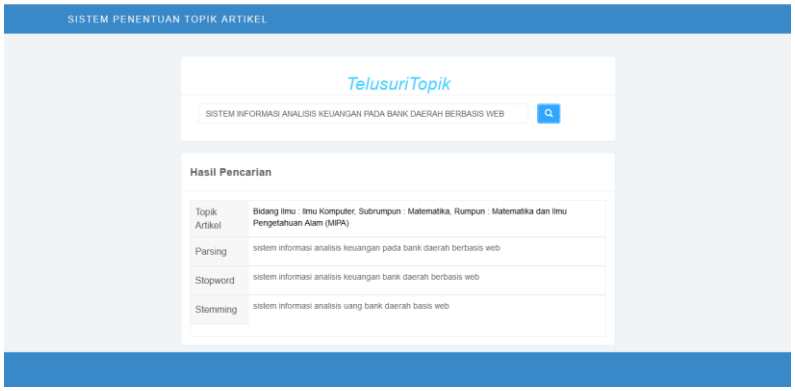
4. IMPLEMENTASI SISTEM

4.1. Implementasi Sistem



Gambar 11 Tampilan *usser*

Gambar 11 menampilkan halaman tampilan *user* dimana terdapat kolom untuk menginputkan judul artikel yang akan ditampilkan topiknya dan menekan tombol *search* untuk menampilkan halaman hasil pencarian topik



Gambar 12 tampilan hasil pencarian

Selanjutnya pada gambar 12 akan menampilkan halaman pencarian dari *query* inputan yang sebelumnya telah dimasukkan. Pada hasil pencarian akan muncul topik artikel yaitu bidang imu, subrumpun ilmu, dan rumpun dari judul artikel yang diintputkan. Kemudian ada beberapa proses yang akan diperlihatkan juga kepada *user* yaitu proses *parsing*, proses penghilangan *stopword*, dan proses *stemming*.

Gambar 13 tampilan input kosong

Gambar 13 menampilkan halaman yang muncul jika *user* melakukan pencarian topik artikel tetapi sebelumnya tidak menginputkan *query* inputan pada kolom yang telah disediakan.

5. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan implementasi sistem penentuan topik artikel menggunakan algoritma *K-NearestNeighbor*, maka dapat disimpulkan bahwa :

1. Hasil penelitian menunjukkan bahwa algoritma yang diterapkan yaitu algoritma *K-Nearest Neighbor* berhasil diimplementasikan pada sistem penentuan topik artikel.
2. Sistem penentuan topik artikel digunakan untuk mengetahui topik sebuah artikel berdasarkan bidang ilmu, subrumpun ilmu, dan rumpun ilmu.

5.2 Saran

Adapun saran dari peneliti tentang penelitian yang telah dilakukan adalah sebagai berikut :

1. Data rumpun ilmu yang diambil adalah data tahun 2012 untuk selanjutnya diharapkan ada pembaharuan data rumpun ilmu.
2. Perlu adanya tambahan pada tampilan *user* yaitu pada hasil pencarian berupa judul-judul artikel yang berkaitan dengan *query* inputan.

DAFTAR PUSTAKA

[1] W. Hakim, "Teknik Menulis Artikel Ilmiah," *Pelatih. Penulisan Prod. Pengetah. KMP PNPM Mandiri Perkota. Wil. 1*, p. 3, 2012.

[2] Y. Samuel, R. Delima, and A. Rachmat, "Implementasi Metode K-Nearest Neighbor dengan Decision Rule untuk Klasifikasi Subtopik Berita," *J. Inform.*, vol. 10, 2015.

[3] A. A. A. Adenowo and B. A. Adenowo, "Software Engineering Methodologies: A Review of the Waterfall Model and Object-Oriented Approach," vol. 4, no. 7, pp. 427–434, 2013.

[4] A. A. Fatahillah and E. Rainarli, "Implementation of Method of Generalized Vector Space Model (GVSM) By Lesk Algorithm on Information Retrieval System," *J. Ilm. dan Inform.*, p. 46, 2012.