**BAB III**

**METODOLOGI PENELITIAN DAN PENGEMBANGAN SISTEM**

Pada bab ini akan dibahas mengenai perancangan dalam membangun sistem. Seperti  
yang dijelaskan pada sebelumnya, dalam proses perancangan membangun sistem  
menggunakan metode prototype. Berikut tahapan yang dilakukan dalam metode prototypeyang diterapkan pada sistem ini.

# **Perencanaan Kebutuhan**

Pada tahap ini dilakukan untuk mendalami seluruh kebutuhan sistem, dan gambaran utama sistem Berikut berapa tahap yang dilakuan pengumpulan kebutuhan :

## **Pengumpulan Data**

Menggunakan jenis data dengan pendekatan kuantitatif, karena penelitian mengutamakan penggunan perhitungan klasifikasi dalam mengklasifikasikan kategori berita.

## **Sumber Data**

Yang digunakan untuk permasalahan ini yaitu :

1. Data Primer

Data primer adalah data yang didapat secara langsung dari sumber asli. Dalam hal ini data yang diperoleh dari scraping portal berita kompas.com, tribunnews.co, detik.com.

1. Data Sekunder

Data sekunder adalah data yang didapat melalui jurnal ilmiah, internet serta laporan penelitian, dan lain-lain yang dapat membantu kelengkapan data dalam landasan teori.

* 1. **Teknik Pengumpulan Data**

Beberapa teknik mengumpulkan data dalam penelitian ini, seperti :

1. Observasi

Metode pendekatan dengan melakukan pengamatan secara langsung pada web portal berita online.

1. Studi Pustaka

Teknik pengumpulan data dengan mengumpulkan informasi dari literatur, internet, jurnal serta sumber lain yang masih ada kaitannya dengan penelitian secara tidak langsung. Dengan adanya pustaka, dapat memperkuat hasil penyelesaian permasalahan.

* + 1. ***Communication***

1. Pengumpulan data

Pengumpulan data terdiri dari jenis dan sumber data, serta cara mengumpulkan data yang akan digunakan. Proses ini juga memuat tentang Analisa kebutuhan sistem dan *user*.

1. Jenis dan sumber data

Jenis data pada penelitian yang dilakukan adalah data kualitatif. Data kualitatif adalah data yang tidak dapat diukur oleh numerik, atau data yang disajikan dalam bentuk verbal. Penelitian yang dilakukan menggunakan data kualitatif dikarenakan data yang akan diambil berupa artikel dengan topik kesehatan, dimana data tersebut tidak dapat diukur dengan numerik.

Sumber data yang akan digunakan dalam penelitian adalah data sekunder, dikarenakan data yang akan digunakan dalam penelitian diambil dari empat buah situs resmi bermuat konten kesehatan yaitu liputan6.com, detik.com, kompas.com, serta klikdokter.com. Hal ini menunjukkan kesesuaian dengan pengertian data sekunder, dimana data sekunder adalah data yang diperoleh dari sumber yang sudah ada. Perolehan sumber data didapatkan dengan proses scrapping.

1. Prosedur pengumpulan data

Pengumpulan data dilakukan dengan melakukan proses scrapping.Proses scrapping sendiri merupakan proses untuk mengambil informasi dari situs di internet. Scrapping bekerja secara otomatis untuk mengumpulkan informasi. Setiap mengunjungi situs, maka setiap link yang ada pada halaman yang dikunjungi akan dikunjungi satu persatu.

Proses pengumpulan data, dimulai dengan mendeklarasikan halaman situs yang akan diambil, yaitu situs liputan6.com, kompas.com, detik.com, dan klikdokter. Web scrapping akan mengunjungi setiap situs, dan mengindeks kumpulan link berita yang ada dalam situs tersebut, satu persatu dikunjungi dan mengambil data judul, nama penulis, waktu penerbitan, halaman URL, serta isi artikel. Hasil dari pengambilan data disimpan didalam database.

Situs dipilih berdasarkan top rank pengunjung pada search engine. Keempat situs terpilih berada dalam Top 50 situs di Indonesia. Hasil ini didapatkan lewat hasil perhitungan kepadatan traffic yang dilakukan oleh alexa.com. Situs ini adalah situs resmi dari amazon.com company yang berfungsi untuk menghitung top sites di setiap negara.

1. Analiss kebutuhan system

Analisis kebutuhan sistem bertujuan untuk mendeskripsikan kebutuhan apa saja yang dilakukan sistem agar sistem dapat berjalan. Untuk memudahkan analisis kebutuhan sistem, kebutuhan sistem dibagi menjadi yaitu, kebutuhan fungsional dan kebutuhan non- fungsional.

1. Kebutuhan fungsional

Kebutuhan fungsional adalah kebutuhan yang akan dikerjakan oleh sistem. Kebutuhan fungsional yang akan dimiliki oleh sistem adalah:

1. Sistem dapat melakukan pencarian kata kunci dan menampilkan artikel yang mempunya hubungan semantik dengan kata kunci

2. Sistem dapat menampilkan ringkasan teks dan ekstraksi keywords secara otomatis bersama dengan hasil pencarian.

3. Sistem dapat mengelola data admin

4. Sistem dapat menambah data dengan implementasi proses scrapping

5. Sistem dapat mengelola daftar sinonim kata

6. Sistem dapat mencatat riwayat hasil pencarian

1. Kebutuhan non-fungsional

1. Analisis kebutuhan perangkat keras

Penelitian dilakukan dengan dukungan perangkat keras, yaitu satu buah unit laptop HP 15 BA004AX dengan spesifikasi sebagai berikut:

a. AMD Quad-Core A10-9600P (2.4 GHz, up to 3.3 GHz, 2 MB cache)

b. Random Access Memory (RAM) 8

c. Hard Disk 1 TB 5400 rpm SATA

d. Resolusi monitor 1366x768

e. AMD Radeon™ R8 M445DX 2GB Dual Graphic

2. Analisis kebutuhan perangkat lunak

Perangkat lunak yang digunakan dalam penelitian adalah sebagai berikut:

a. Windows sebagai operating system

b. Python sebagai Bahasa pemrograman

c. Flask sebagai framework pemograman python

d. Sublime dan Spyder sebagai tools editor program

e. CMD digunakan untuk membuat lingkungan executing program

f. XAMPP Control Panel

g. MySQL database

3. Analisis kebutuhan pengguna

Kebutuhan pengguna dalam sistem ini adalah sebagai berikut:

a. Pengguna dapat mendapatkan informasi berita kesehatan serta mendapatkan intisari bahasan artikel hasil pencarian

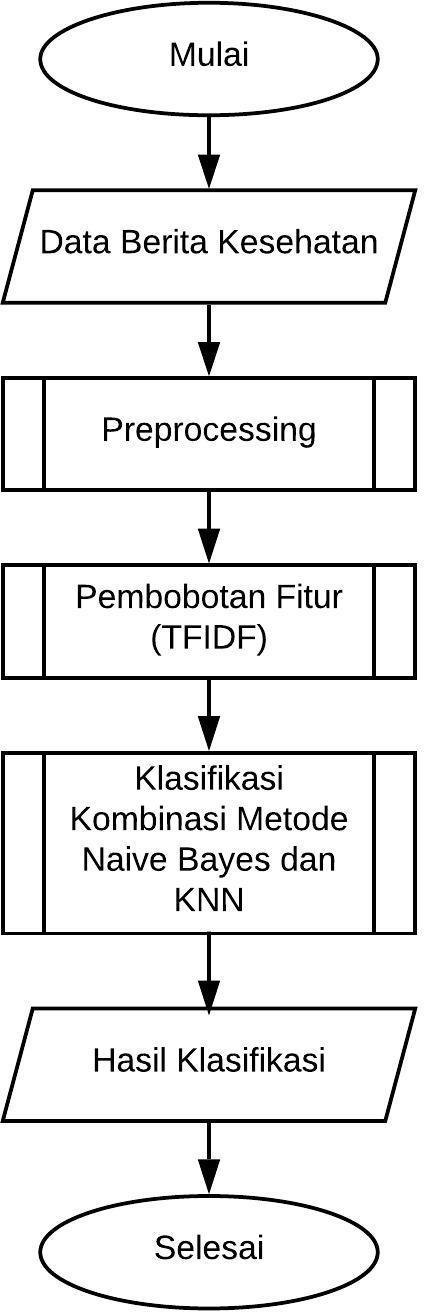
* + 1. ***Quick Plan***

Perancangan sistem akan digambarkan dengan menggunakan diagram alir atau *flowchart* agar alur sistem yang digambarkan dapat terlihat dengan jelas. Pada subbab ini juga akan dilakukan analisis terhadap langkah-langkah preprocessing, pembobotan fitur, dan proses klasifikasi.

* + - 1. *Flowchart* Identifikasi

Pada tahap ini dijelaskan proses identifikasi artikel. Dimulai dari inputan *user*, kemudian dilakukan *preprocessing text*, setelah itu dilakukan proses klasifikasi menggunakan metode kombinasi *Naïve Bayes* dan *K-Nearest Neighbor*. Hasilnya akan

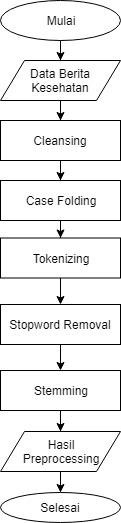
diketahui kelas dari suatu artikel inputan *user*, apakah termasuk kelas fakta atau *hoax*. Gambar 3.1 merupakan *flowchart* identifikasi.



**Gambar 3.1** Gambar *Flowchart* Identifikasi

* + - 1. *Flowchart* Subproses *Preprocessing*

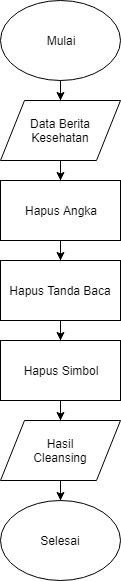
Subproses dari proses *preprocessing* meliputi proses *cleansing, case folding, tokenizing, stopword removal,* dan *stemming.* Proses ini merupakan proses untuk menjadikan data teks yang belum terolah, menjadi siap pakai untuk proses klasifikasi teks. Gambar 3.2 merupakan *flowchart preprocessing*.



**Gambar 3.2** Gambar *Flowchart* Preprocessing

* + - 1. *Flowchart* Subproses *Cleansing*

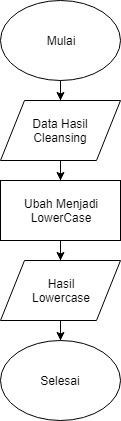
Subproses *cleansing* merupakan tahap pembersihan teks dari gangguan yang dapat menyebabkan *noise*. Semua bentuk selain huruf ‘a’ sampai ‘z’ dibersihkan, termasuk diantaranya angka, tanda baca, simbol, *hashtag*, dan url akan dihapus karena tidak memiliki pengaruh apapun terhadap perhitungan klasifikasi. Gambar 3.3 merupakan *flowchart cleansing*.



**Gambar 3.3** Gambar *Flowchart* proses *cleansing*

* + - 1. *Flowchart* Subproses *Case Folding*

Subproses *case folding* merupakan proses pengubahan huruf dalam dokumen menjadi satu bentuk, misalnya huruf kapital menjadi huruf kecil dan sebaliknya. Pada penelitian ini akan menyeragamkan semua huruf dalam dokumen menjadi huruf kecil atau biasa disebut *lowercase*. Gambar 3.4 merupakan *flowchart case folding*.



**Gambar 3.4** Gambar *Flowchart* proses *Case Folding*

* + - 1. *Flowchart* Subproses *Tokenizing*

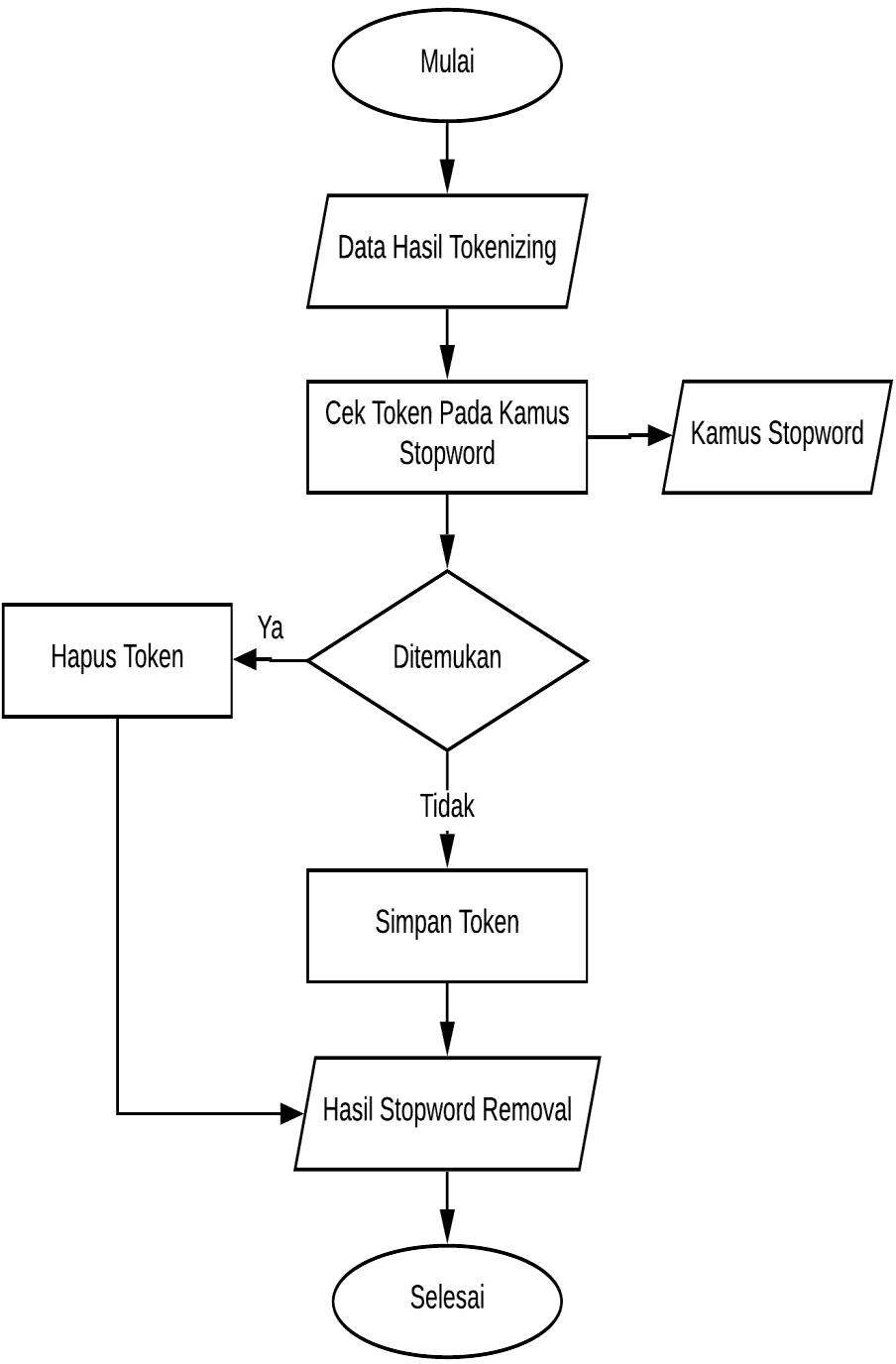
Subproses *tokenizing* merupakan proses pemotongan string masukan berdasarkan tiap- tiap kata yang menyusunnya. Pada umumnya setiap kata terindentifikasi dengan kata yang lain oleh karakter spasi, sehingga proses *tokenization* mengendalikan karakter spasi pada dokumen untuk pemisahan kata. Gambar 3.5 merupakan *flowchart tokenizing*.



**Gambar 3.5** Gambar *Flowchart* proses *Tokenizing*

* + - 1. *Flowchart* Subproses *Stopword Removal*

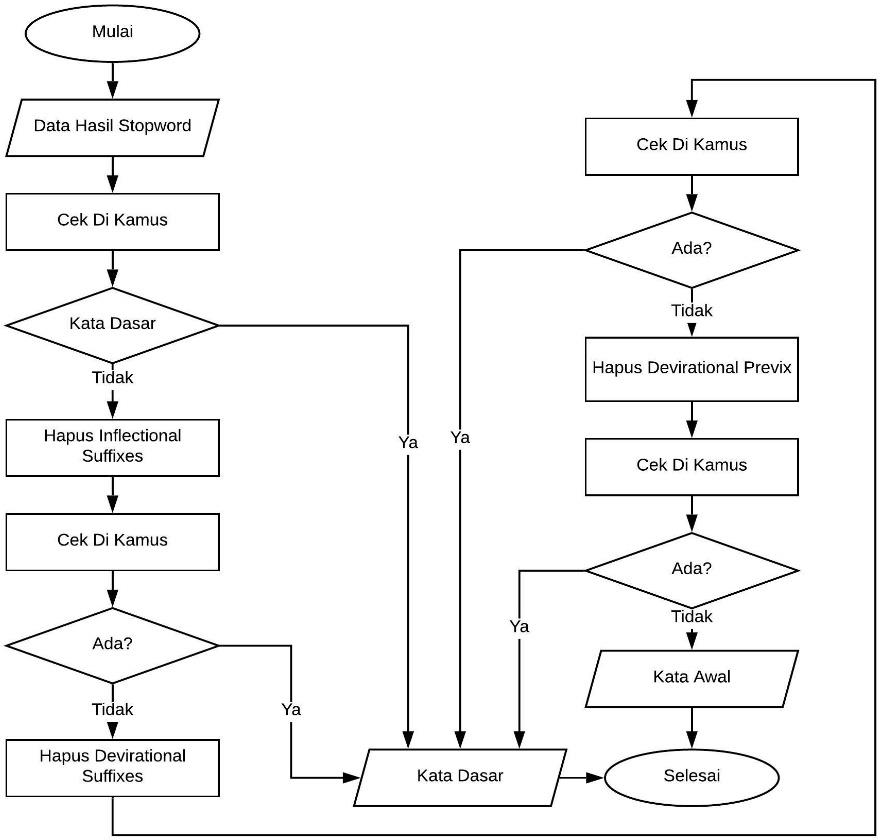
Subproses *stopword removal* merupakan proses pembuangan daftar kata yang berupa *stopword*. Proses ini adalah proses pengambilan kata-kata yang dianggap penting atau mempunyai makna saja. Gambar 3.6 merupakan *flowchart stopword removal*.



**Gambar 3.6** Gambar *Flowchart* proses *Stopword Removal*

* + - 1. *Flowchart* Subproses *Stemming*

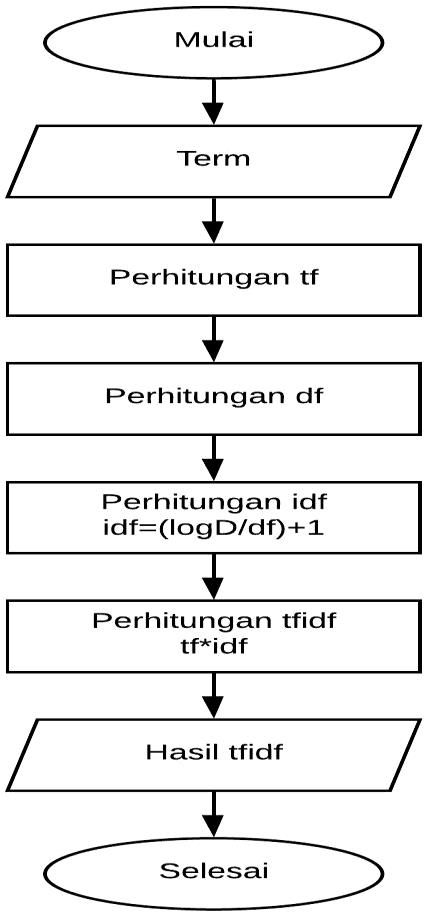
Subproses *stemming* merupakan proses untuk mendapatkan *root* atau *stem* atau kata dasar dari suatu kata dalam kalimat. Proses *stemming* dilakukan dengan cara memisahkan masing-masing kata dari kata dasar dan imbuhannya baik awalan (*prefiks*) maupun akhiran (*sufiks*). Salah satu tujuan dilakukan proses *stemming* yaitu untuk meningkatkan efisiensi dengan cara memilah isi dokumen menjadi unit-unit kecil yang akan menjadi penciri misalnya berupa kata. Gambar 3.7 merupakan *flowchart stemming*.



**Gambar 3.7** Gambar *Flowchart* proses Stemming

* + - 1. *Flowchart* Subproses pemboboran fitur TFIDF

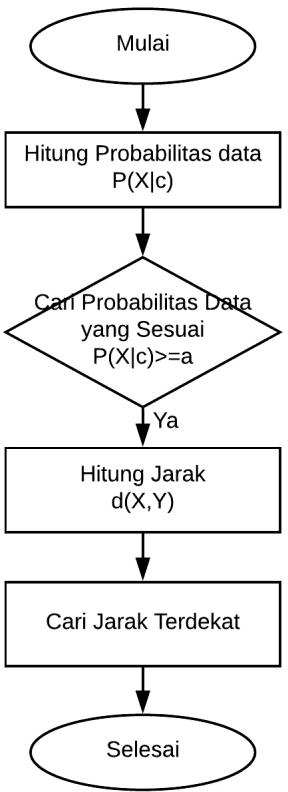
Subproses perhitungan TFIDF adalah proses perhitungan bobot tiap kata. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu *term frequency* (TF) dan *invers document frequency* (IDF). Gambar 3.8 merupakan *flowchart* pembobotan fitur *TFIDF*.



**Gambar 3.8** Gambar *Flowchart* proses TFIDF

* + - 1. *Flowchart* Subproses klasifikasi metode kombinasi Naïve Bayes dan KNN Subproses klasifikasi dengan metode kombinasi Naïve Bayes dan KNN adalah proses

identifikasi kelas dari suatu dokumen atau artikel. Metode ini akan mengklasifikasikan suatu dokumen uji dengan cara mencari nilai probabilitas dari suatu dokumen latih yang lebih besar dari α (pada penelitian (Sari, Ernawati, & Pranowo, 2015) nilai α adalah 0.5). Data latih yang lolos kemudian akan dibandingkan dengan data uji untuk dicari kedekatan jaraknya menggunakan Cosines Similarity. Jarak dengan nilai terbesar berarti dokumen uji paling dekat dengan dokumen latih tersebut. Gambar 3.9 merupakan *flowchart* klasifikasi dengan metode kombinasi *Naïve Bayes* dan KNN.



**Gambar 3.9** Gambar *Flowchart* klasifikasi metode kombinasi *Naïve Bayes* dan KNN

* + 1. **Analisis Teks Preprocessing**

Teks Preprocessing merupakan langkah awal dalam mempersiapkan data agar data dapat diolah pada proses klasifikasi. *Preprocessing* berfungsi untuk membersihkan data dari karakter-karakter yang tidak diperlukan yang dapat mengganggu proses klasifikasi dan juga membuat kata menjadi seragam agar dapat memudahkan dalam melakukan pengolahan selanjutnya. Tahapan-tahapan dalam teks *preprocessing* antara lain:

1. *Cleansing*

*Cleansing* merupakan proses pembersihan teks dari karakter yang tidak diperlukan yang menyebabkan adanya noise. Untuk karakter yang dihapuskan yaitu selain huruf ‘a’ sampai ‘z’ seperti symbol, tanda baca, dan juga url yang tidak memiliki arti sehingga dapat

menggangu proses klasifikasi. Untuk menghilangkan karakter tersebut akan diubah menjadi karakter spasi.

1. *Case Folding*

*Case Folding* merupakan tahap pengubahan huruf dalam dokumen menjadi seragam, sehingga harus dibuat menjadi semua huruf kecil saja atau huruf besar semua. Untuk karakter selain huruf akan diabaikan dan disebut sebagai delimeter.

1. *Tokenizing*

*Tokenizing* merupakan tahapan untuk membagi dokumen teks yang dapat berupa kalimat atau paragraf menjadi token-token atau bagian-bagian tertentu yang dipisahkan oleh karakter spasi yang digunakan untuk menghitung frekuensi pada setiap kata.

1. *Stopword Removal*

*Stopword Removal* merupakan tahap penghapusan kata yang tidak mempunyai makna ataupun informasi yang bisa membantu dalam proses klasifikasi.

1. *Stemming*

Stemming merupakan tahap untuk menghilangkan imbuhan suatu kata yang berupa awalan dan akhiran sehingga menjadi sebuah kata dasar.

* + 1. **Analisis Pembobotan Fitur dengan TF-IDF**

Pada tahap pelatihan ini dilakukan proses permbobotan atau ekstraksi fitur pada suatu kata yaitu dengan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Berikut adalah langkah-langkah dalam tahapan ekstraksi fitur ini:

Dokumen Latih:

Dokumen 1: Tim peneliti, yang dipimpin oleh Profesor Mark Wilcox dari *School of Medicine*, membuktikan tangan terkontaminasi dengan jenis berbahaya dari bakteri yang

disebut Lactobacillus, yang tidak biasanya ditemukan di kamar mandi umum. Deteksi berikutnya dari Lactobacillus di udara membuktikan bahwa bakteri tersebut berasal dari tangan selama pengeringan. Para ahli mengumpulkan sampel udara di sekitar pengering tangan dan juga pada jarak satu dan dua meter. (Fakta)

Dokumen 2: Tidak ada bukti medis untuk mendukung ‘batuk CPR’ saat mengalami serangan jantung ketika Anda sendirian. Mitos ‘batuk CPR’ telah beredar di internet sejak, terutama di situs media sosial seperti Facebook. Jika Anda menemukannya pesan itu, mohon jangan menyebarkannya dan beritahu bahwa saran itu tidak benar. (Fakta)

Dokumen 3: Menyalakan AC saat pertama masuk mobil dengan jendela tertutup akan mengandung 400-800 mg Benzene. Jika parkir di luar ruangan dibawah sinar matahari pada temperatur 15 derajat ke atas, tingkat Benzene naik hingga 2000-4000 mg, 40x dari tingkat Benzene yang bisa ditoleransi. Orang-orang yang sewaktu masuk ke dalam mobil, tidak membuka jendela sama sekali, berakibat akan menghirup banyak racun yang dengan sangat cepat masuk kedalam tubuh anda. (Hoax)

Dokumen 4: HATI-HATI ADA BESI TIPIS DI DALAM OBAT. SEBARKAN!!!

Bila kita bakal minum obat, sebelum meminumnya alangkah baiknya kita belah jadi 2 ataupun bila butuh kita belah jadi 4/sekaligus kita remuk gunakan sendok teh. (Hoax)

Hasil Preprocessing:

Dokumen 1: tim teliti pimpin profesor mark wilcox school of medicine bukti tangan kontaminasi jenis bahaya bakteri sebut lactobacillus temu kamar mandi umum deteksi lactobacillus udara bukti bakteri tangan kering ahli kumpul sampel udara sekitar kering tangan jarak meter

Dokumen 2: bukti medis dukung batuk cpr alami serang jantung mitos batuk cpr edar internet situs media sosial facebook temu pesan sebar beritahu saran

Dokumen 3: nyala ac mobil jendela tutup kandung mg benzene parkir luar ruang bawah sinar matahari temperatur derajat tingkat benzene mg tingkat benzene toleransi waktu mobil buka jendela hirup racun cepat tubuh

Dokumen 4: hati hati besi tipis obat sebar minum obat minum belah butuh belah remuk sendok teh

1. Perhitungan TF Dokumen Uji:

Para ilmuwan dari University of Leeds telah menemukan bahwa mesin pengering tangan atau ‘jet-air dryer’ dapat menyebarkan bakteri di toilet.

Hasil Preprocessing: ilmuwan university of leeds temu mesin kering tangan jet air dryer sebar bakteri toilet

Frekuensi dari dokumen-dokumen diatas dapat di perhatikan pada tabel 3.1

**Tabel 3.1** Tabel contoh proses perhitungan tf

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NO | Term | TF | | | |
| D1 | D2 | D3 | D4 |
| 1 | Tim | 1 | 0 | 0 | 0 |
| 2 | Teliti | 1 | 0 | 0 | 0 |
| 3 | Pimpin | 1 | 0 | 0 | 0 |
| 4 | Professor | 1 | 0 | 0 | 0 |
| 5 | Mark | 1 | 0 | 0 | 0 |
| 6 | Wilcox | 1 | 0 | 0 | 0 |
| 7 | school | 1 | 0 | 0 | 0 |
| 8 | of | 1 | 0 | 0 | 0 |
| 9 | medicine | 1 | 0 | 0 | 0 |
| 10 | bukti | 2 | 1 | 0 | 0 |
| 11 | tangan | 3 | 0 | 0 | 0 |
| 12 | kontaminasi | 1 | 0 | 0 | 0 |
| 13 | jenis | 1 | 0 | 0 | 0 |
| 14 | bahaya | 1 | 0 | 0 | 0 |
| 15 | bakteri | 2 | 0 | 0 | 0 |
| 16 | sebut | 1 | 0 | 0 | 0 |
| 17 | lactobacillus | 2 | 0 | 0 | 0 |

**Tabel 3.2** Lanjutan tabel contoh proses perhitungan tf

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 18 | temu | 1 | 1 | 0 | 0 |
| 19 | kamar | 1 | 0 | 0 | 0 |
| 20 | mandi | 1 | 0 | 0 | 0 |
| 21 | umum | 1 | 0 | 0 | 0 |
| 22 | deteksi | 1 | 0 | 0 | 0 |
| 23 | udara | 2 | 0 | 0 | 0 |
| 24 | kering | 2 | 0 | 0 | 0 |
| 25 | ahli | 1 | 0 | 0 | 0 |
| 26 | kumpul | 1 | 0 | 0 | 0 |
| 27 | sampel | 1 | 0 | 0 | 0 |
| 28 | sekitar | 1 | 0 | 0 | 0 |
| 29 | jarak | 1 | 0 | 0 | 0 |
| 30 | meter | 1 | 0 | 0 | 0 |
| 31 | medis | 0 | 1 | 0 | 0 |
| 32 | dukung | 0 | 1 | 0 | 0 |
| 33 | batuk | 0 | 2 | 0 | 0 |
| 34 | cpr | 0 | 2 | 0 | 0 |
| 35 | alami | 0 | 1 | 0 | 0 |
| 36 | serang | 0 | 1 | 0 | 0 |
| 37 | jantung | 0 | 1 | 0 | 0 |
| 38 | mitos | 0 | 1 | 0 | 0 |
| 39 | edar | 0 | 1 | 0 | 0 |
| 40 | internet | 0 | 1 | 0 | 0 |
| 41 | situs | 0 | 1 | 0 | 0 |
| 42 | media | 0 | 1 | 0 | 0 |
| 43 | sosial | 0 | 1 | 0 | 0 |
| 44 | facebook | 0 | 1 | 0 | 0 |
| 45 | pesan | 0 | 1 | 0 | 0 |
| 46 | sebar | 0 | 1 | 0 | 1 |
| 47 | beritahu | 0 | 1 | 0 | 0 |
| 48 | saran | 0 | 1 | 0 | 0 |
| 49 | nyala | 0 | 0 | 1 | 0 |
| 50 | ac | 0 | 0 | 1 | 0 |
| 51 | mobil | 0 | 0 | 2 | 0 |
| 52 | jendela | 0 | 0 | 2 | 0 |
| 53 | tutup | 0 | 0 | 1 | 0 |
| 54 | kandung | 0 | 0 | 1 | 0 |
| 55 | mg | 0 | 0 | 2 | 0 |
| 56 | benzene | 0 | 0 | 3 | 0 |
| 57 | parkir | 0 | 0 | 1 | 0 |
| 58 | luar | 0 | 0 | 1 | 0 |
| 59 | ruang | 0 | 0 | 1 | 0 |
| 60 | bawah | 0 | 0 | 1 | 0 |
| 61 | sinar | 0 | 0 | 1 | 0 |
| 62 | matahari | 0 | 0 | 1 | 0 |
| 63 | temperatur | 0 | 0 | 1 | 0 |
| 64 | derajat | 0 | 0 | 1 | 0 |
| 65 | tingkat | 0 | 0 | 2 | 0 |
| 66 | toleransi | 0 | 0 | 1 | 0 |
| 67 | waktu | 0 | 0 | 1 | 0 |
| 68 | buka | 0 | 0 | 1 | 0 |
| 69 | hirup | 0 | 0 | 1 | 0 |

**Tabel 3.3** Lanjutan tabel contoh proses perhitungan tf

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 70 | racun | 0 | 0 | 1 | 0 |
| 71 | cepat | 0 | 0 | 1 | 0 |
| 72 | tubuh | 0 | 0 | 1 | 0 |
| 73 | hati | 0 | 0 | 0 | 2 |
| 74 | besi | 0 | 0 | 0 | 1 |
| 75 | tipis | 0 | 0 | 0 | 1 |
| 76 | obat | 0 | 0 | 0 | 2 |
| 77 | minum | 0 | 0 | 0 | 2 |
| 78 | belah | 0 | 0 | 0 | 2 |
| 79 | butuh | 0 | 0 | 0 | 1 |
| 80 | remuk | 0 | 0 | 0 | 1 |
| 81 | sendok | 0 | 0 | 0 | 1 |
| 82 | teh | 0 | 0 | 0 | 1 |

1. Perhitungan TF-IDF

Untuk perhitungan TFIDFnya menggunakan *library* TFIDF *Scikit-learn*. Sehingga perhitungan TFIDF adalah sebagai berikut:

D1 : *tf.idf(ilmuwan).D1* = 1 x 2.386= 2.386

D2 : *tf.idf(ilmuwan).D2* = 0 x 2.386= 0

D3 : *tf.idf(ilmuwan).D3* = 0 x 2.386= 0

D4 : *tf.idf(ilmuwan).D4* = 0 x 2.386= 0

Perhitungan dari setiap *term* dapat dilihat pada table 3.4

**Tabel 3.4** Tabel Contoh proses pehitungan tfidf

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NO | Term | df | D/df | IDF | W=(tf\*IDF) | | | |
| D1 | D2 | D3 | D4 |
| 1 | tim | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 2 | teliti | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 3 | pimpin | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 4 | profesor | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 5 | mark | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 6 | wilcox | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 7 | school | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 8 | of | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 9 | medicine | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 10 | bukti | 2 | 2 | 1.693 | 3.386 | 1.693 | 0 | 0 |
| 11 | tangan | 2 | 2 | 1.693 | 5.079 | 0 | 0 | 0 |
| 12 | kontaminasi | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 13 | jenis | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 14 | bahaya | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 15 | bakteri | 2 | 2 | 1.693 | 3.386 | 0 | 0 | 0 |

**Tabel 3.5** Lanjutan Tabel contoh proses pehitungan tfidf

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 16 | sebut | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 17 | lactobacillus | 1 | 4 | 2.386 | 4.772 | 0 | 0 | 0 |
| 18 | temu | 3 | 1.33 | 1.287 | 1.287 | 1.287 | 0 | 1.287 |
| 19 | kamar | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 20 | mandi | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 21 | umum | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 22 | deteksi | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 23 | udara | 1 | 4 | 2.386 | 4.772 | 0 | 0 | 0 |
| 24 | kering | 2 | 2 | 1.693 | 3.386 | 0 | 0 | 0 |
| 25 | ahli | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 26 | kumpul | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 27 | sampel | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 28 | sekitar | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 29 | jarak | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 30 | meter | 1 | 4 | 2.386 | 2.386 | 0 | 0 | 0 |
| 31 | medis | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 32 | dukung | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 33 | batuk | 1 | 4 | 2.386 | 0 | 4.772 | 0 | 0 |
| 34 | cpr | 1 | 4 | 2.386 | 0 | 4.772 | 0 | 0 |
| 35 | alami | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 36 | serang | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 37 | jantung | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 38 | mitos | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 39 | edar | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 40 | internet | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 41 | situs | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 42 | media | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 43 | sosial | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 44 | facebook | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 45 | pesan | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 46 | sebar | 3 | 1.33 | 1.287 | 0 | 1.287 | 0 | 1.287 |
| 47 | beritahu | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 48 | saran | 1 | 4 | 2.386 | 0 | 2.386 | 0 | 0 |
| 49 | nyala | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 50 | ac | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 51 | mobil | 1 | 4 | 2.386 | 0 | 0 | 4.772 | 0 |
| 52 | jendela | 1 | 4 | 2.386 | 0 | 0 | 4.772 | 0 |
| 53 | tutup | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 54 | kandung | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 55 | mg | 1 | 4 | 2.386 | 0 | 0 | 4.772 | 0 |
| 56 | benzene | 1 | 4 | 2.386 | 0 | 0 | 7.158 | 0 |
| 57 | parkir | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 58 | luar | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 59 | ruang | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 60 | bawah | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 61 | sinar | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 62 | matahari | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 63 | temperatur | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 64 | derajat | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 65 | tingkat | 1 | 4 | 2.386 | 0 | 0 | 4.772 | 0 |
| 66 | toleransi | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 67 | waktu | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |

**Tabel 3.6** Lanjutan contoh proses pehitungan tfidf

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 68 | buka | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 69 | hirup | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 70 | racun | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 71 | cepat | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 72 | tubuh | 1 | 4 | 2.386 | 0 | 0 | 2.386 | 0 |
| 73 | hati | 1 | 4 | 2.386 | 0 | 0 | 0 | 4.772 |
| 74 | besi | 1 | 4 | 2.386 | 0 | 0 | 0 | 2.386 |
| 75 | tipis | 1 | 4 | 2.386 | 0 | 0 | 0 | 2.386 |
| 76 | obat | 1 | 4 | 2.386 | 0 | 0 | 0 | 4.772 |
| 77 | minum | 1 | 4 | 2.386 | 0 | 0 | 0 | 4.772 |
| 78 | belah | 1 | 4 | 2.386 | 0 | 0 | 0 | 4.772 |
| 79 | butuh | 1 | 4 | 2.386 | 0 | 0 | 0 | 2.386 |
| 80 | remuk | 1 | 4 | 2.386 | 0 | 0 | 0 | 2.386 |
| 81 | sendok | 1 | 4 | 2.386 | 0 | 0 | 0 | 2.386 |
| 82 | teh | 1 | 4 | 2.386 | 0 | 0 | 0 | 2.386 |

* + 1. **Analisa Penerapan Naïve Bayes**

Pada proses Naïve Bayes akan dilakukan perhitungan dengan mencari nilai probabilitas data yang akan diklasifikasikan pada masing-masing kelas P(x|c) dengan menggunakan hasil dari perhitungan tf-idf yang terdapat pada tabel. Langkah-langkah dari penerapan metode sebagai berikut:

P(Fakta) = 2/4 = 0.5 P(Hoax) = 2/4 = 0.5

Pertama hitung nilai probabilitas setiap term sesuai pada kelasnya masing-masing dengan cara membandingkan dengan data latih yang telah memiliki label kelas. Untuk hasil probabilitas dapat dilihat pada tabel 3.7

**Tabel 3.7** Tabel term pada kelas

|  |  |  |  |
| --- | --- | --- | --- |
| NO | Term | Kelas | |
| Fakta | hoax |
| 1 | Ilmuwan | 0 | 0 |
| 2 | University | 0 | 0 |
| 3 | Of | 0 | 0 |
| 4 | Leeds | 0 | 0 |
| 5 | Temu | 2 | 0 |
| 6 | Mesin | 0 | 0 |
| 7 | Kering | 2 | 0 |
| 8 | Tangan | 3 | 0 |
| 9 | Jet | 0 | 0 |

**Tabel 3.8** Lanjutan tabel term pada kelas

|  |  |  |  |
| --- | --- | --- | --- |
| 10 | Air | 0 | 0 |
| 11 | Dryer | 0 | 0 |
| 12 | Sebar | 1 | 1 |
| 13 | Bakteri | 2 | 0 |
| 14 | Toilet | 0 | 0 |

Dari hasil probabilitas tiap kelas tersebut dihitung sesuai dengan jumlah term yang sesuai dengan probabilitasnya. Contoh perhitungan dari probabilitas kelas yaitu: P(temu|fakta): 2𝑥0.5 = 1

P(temu|hoax): 0𝑥0.5 = 0

Probabilitas dihitung seperti contoh berikut sesuai dengan jumlah term dan kelas dari tiap- tiap term. Untuk hasil perhitungan dapat dilihat pada tabel 3.9

**Tabel 3.9** Tabel probabilitas term pada kelas

|  |  |  |  |
| --- | --- | --- | --- |
| NO | Term | Kelas | |
| Fakta | hoax |
| 1 | temu | 1 | 0 |
| 2 | kering | 1 | 0 |
| 3 | tangan | 1.5 | 0 |
| 4 | sebar | 0.5 | 0.5 |
| 5 | bakteri | 1 | 0 |

Ketika term telah memiliki nilai pada setiap kelasnya, lalu jumlahkan nilai probabilitas term sesuai dengan kata yang terdapat pada dokumen latih.

Dokumen 1: temu(1) + kering(2) + tangan(3) + bakteri(2) = 1\*1 + 2\*1 + 3\*1.5 + 2\*1 = 9.5 Dokumen 2: temu(1) + sebar(1) = 1\*1 + 1\*0.5 = 1.5

Dokumen 3: 0

Dokumen 4: sebar(1) = 1\*0.5 = 0.5

Ketika semua dokumen telah memiliki nilai probabilitas dari perhitungan menggunakan naïve bayes, maka ditentukan nilai α terlebih dahulu. Nilai α ini berfungsi sebagai ambang batas minimal dari nilai probabilitas tiap dokumen. Apabila nilai probabilitasnya lebih besar atau sama dengan dari nilai α maka dokumen-dokumen tersebut diambil untuk dilakukan

perhitungan selanjutnya, sedangkan untuk nilai yang tidak memenuhi tidak akan digunakan untuk proses perhitungan selanjutnya. Misalkan nilai ditentukan nilai α = 0.5 maka dokumen yang masuk pada perhitungan selanjutnya yaitu Dokumen 1, Dokumen 2 dan Dokumen 4.