

## **BAB 4**

### **PERANCANGAN**

Bab berikut analisis dan perancangan yang membahas pada sistem “Penerapan Klasifikasi *Tweets* Pada Berita Twitter Menggunakan Metode *K-Nearest Neighbor* dan *Query Expansion* Berbasis *Distributional Semantic*”.

#### **4.1 Lingkungan Perancangan dan Implementasi**

Analisis kebutuhan digunakan untuk mengidentifikasi apa saja yang diperlukan dalam pembuatan sistem ini. Kebutuhan-kebutuhan tersebut antara lain:

1. Perangkat keras (*hardware*)
  - Laptop Lenovo intel pentium dual core
  - RAM 2 GB
  - DDR3
  - VGA 1 GB
2. Perangkat lunak (*software*)
  - Microsoft windows 8.1 sebagai sistem operasi yang digunakan
  - JAVA adalah program dasar dari sistem
  - MySql adalah sebagai database

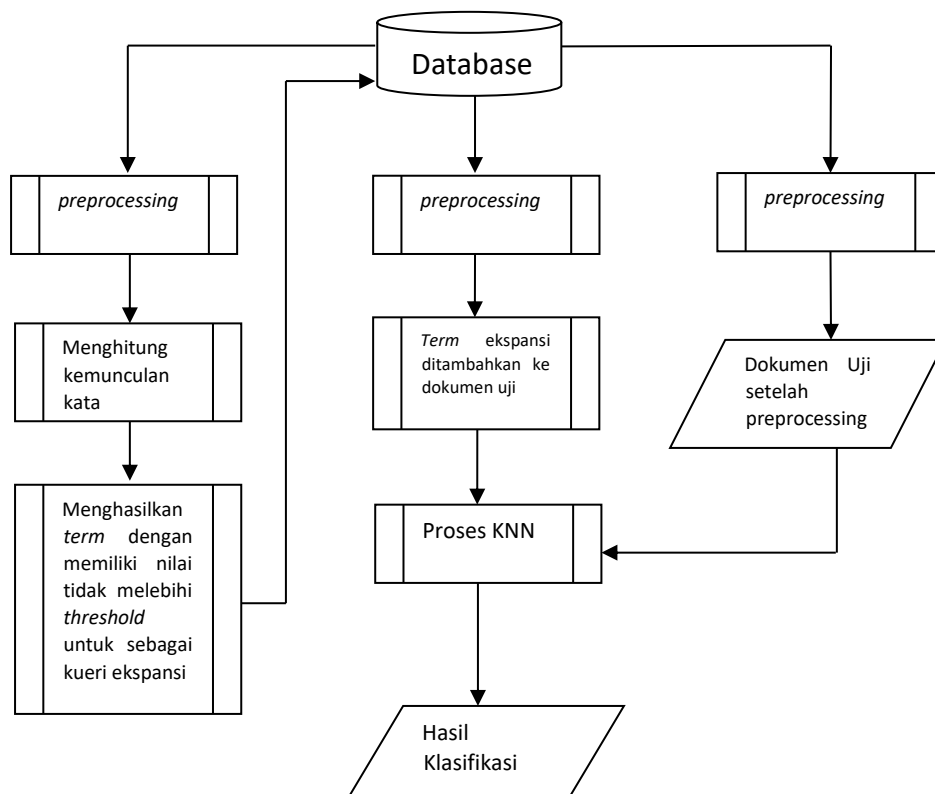
#### **4.2 Perancangan Perangkat Lunak**

Pada perancangan perangkat sistem yang akan dibangun oleh penulis, meliputi deskripsi umum alur sistem, perancangan sistem manajemen data, dan perincian perancangan diagram alir sistem.

##### **4.2.1 Deskripsi Sistem Secara Umum**

Alur kerja sistem dijelaskan pada Gambar 4.1. Prinsip kerja sistem yang akan dibangun bertujuan menghasilkan klasifikasi teks namun lebih rinci lagi bahwa klasifikasi tersebut digunakan untuk menentukan kedekatan kata yang dapat diambil dari data atau kamus berita yang tersedia pada dataset. Proses ekspansi

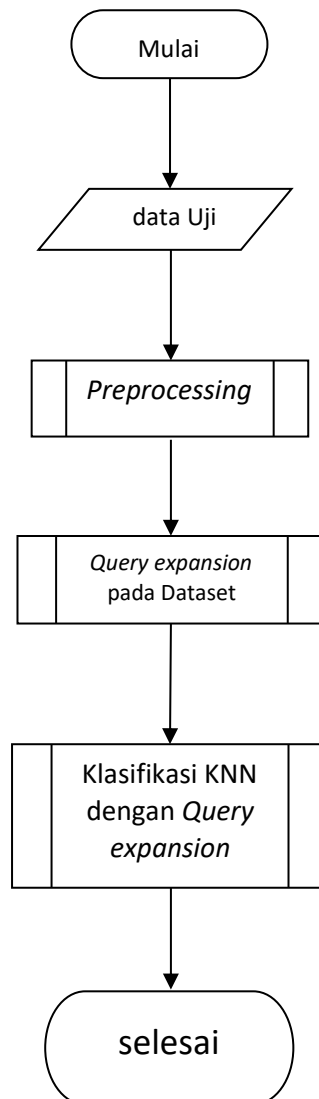
kata dilakukan dengan menghitung nilai dan jumlah frekuensi kemunculan kata untuk pemilihan kata yang akan digunakan sebagai penambahan *term* yang diambil dengan tidak melebihi nilai batas maksimal *threshold*. *Data set*, data uji, dan data latih sebelumnya dilakukan proses preprocessing agar mendapatkan teks atau *term* yang sudah dihilangkan tanda baca, imbuhan, dan kata yang dianggap tidak penting oleh sistem. Hal ini berguna agar hasil yang diperoleh nanti akan mempunyai tingkat akurasi yang lebih baik. Proses ekspansi kata hanya dilakukan pada data uji yang berupa *tweets*, sedangkan kamus berita atau dataset digunakan untuk membantu proses penambahan kata yang terdekat dari setiap kata pada *tweets*. Setelah proses tersebut, berikutnya adalah proses klasifikasi untuk menentukan bahwa *tweets* tersebut termasuk pada konten yang mana. *Dataset* berita yang digunakan berasal dari portal berita detik.com, serta data latih dan data uji berasal dari *tweets* akun *Twitter* Kompas dan detik.



**Gambar 4.1 Diagram Alur Kerja Sistem Secara Umum**

#### 4.2.2 Perancangan Diagram alir sistem

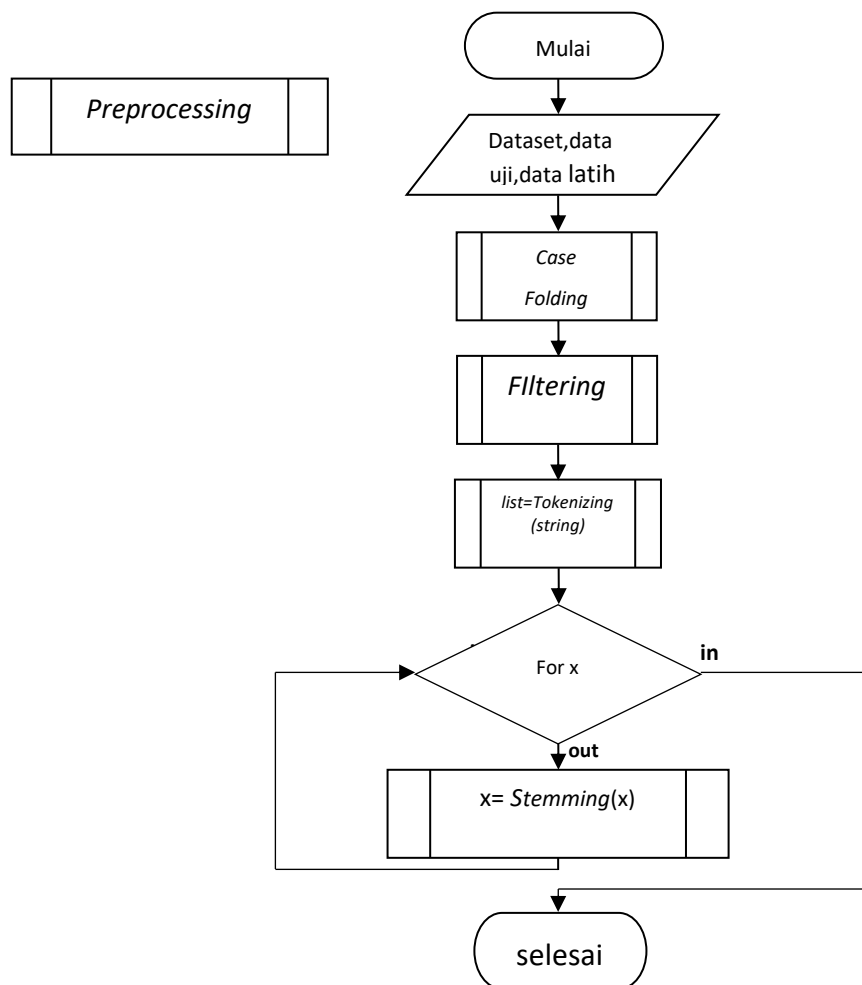
Perancangan sistem pada diagram alir sistem terdapat tiga proses utama yaitu, *preprocessing*, proses algoritme *query expansion*, serta klasifikasi K-Nearest Neighbor. Dalam ketiga tahap tersebut masih terdapat beberapa tahapan lagi di dalamnya untuk lebih mendetailkan. Proses keseluruhannya adalah Gambar 4.2.



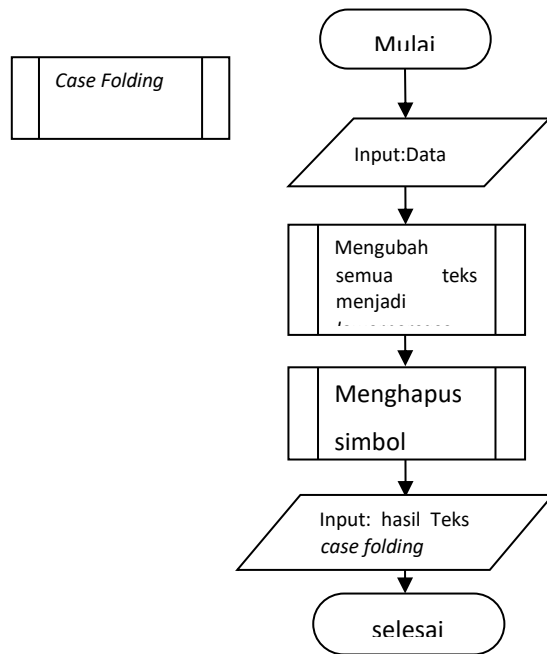
**Gambar 4.2 Diagram Alir Sistem**

Awalnya sistem akan mengambil data yang telah dipilih dan akan di ujikan dalam tahap klasifikasi. Proses berikutnya data uji yang telah dipilih, sistem akan melakukan tahap *preprocessing*, yang berguna untuk menghilangkan kata

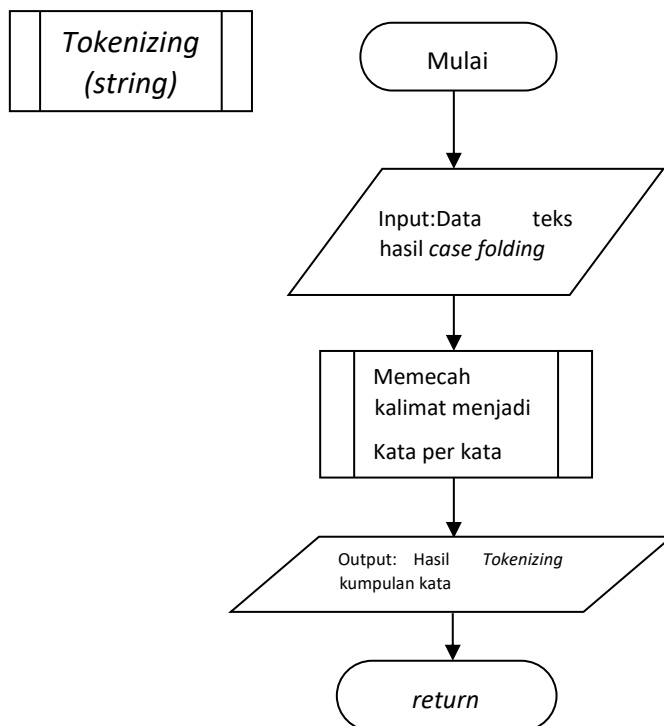
imbuhan, tambahan, kata sambung, tanda baca, dan kata yang berganda agar memiliki kata dasar yang digunakan sebagai *term* untuk diproses oleh sistem. Setelah itu, tahap *query expansion* atau ekspansi kata dilakukan untuk mencari kedekatan kata pada setiap *term* pada *tweets* berdasarkan kamus berita dengan algoritme *query expansion*. Setelah mengalami proses ekspansi kata dan penambahan kata, tahap berikutnya yaitu melakukan klasifikasi data berupa *tweets* sebagai data latih dari *Twitter*. Proses *preprocessing* dapat dijelaskan pada Gambar 4.3, 4.4, 4.5, 4.6, 4.7.



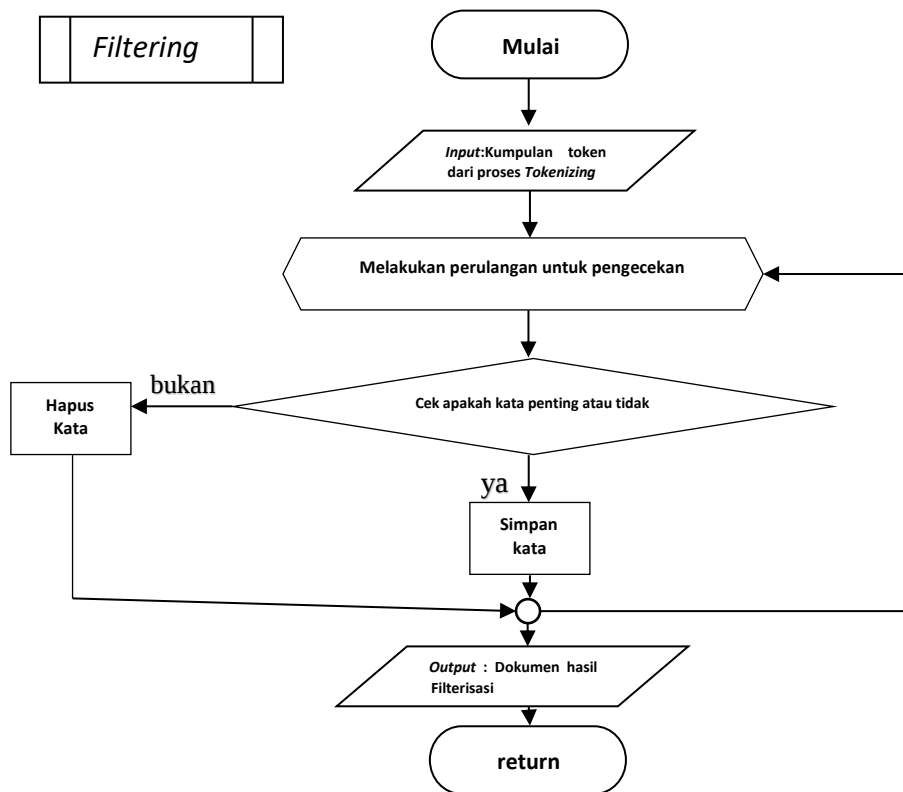
**Gambar 4.3 Diagram Alir *Preprocessing***



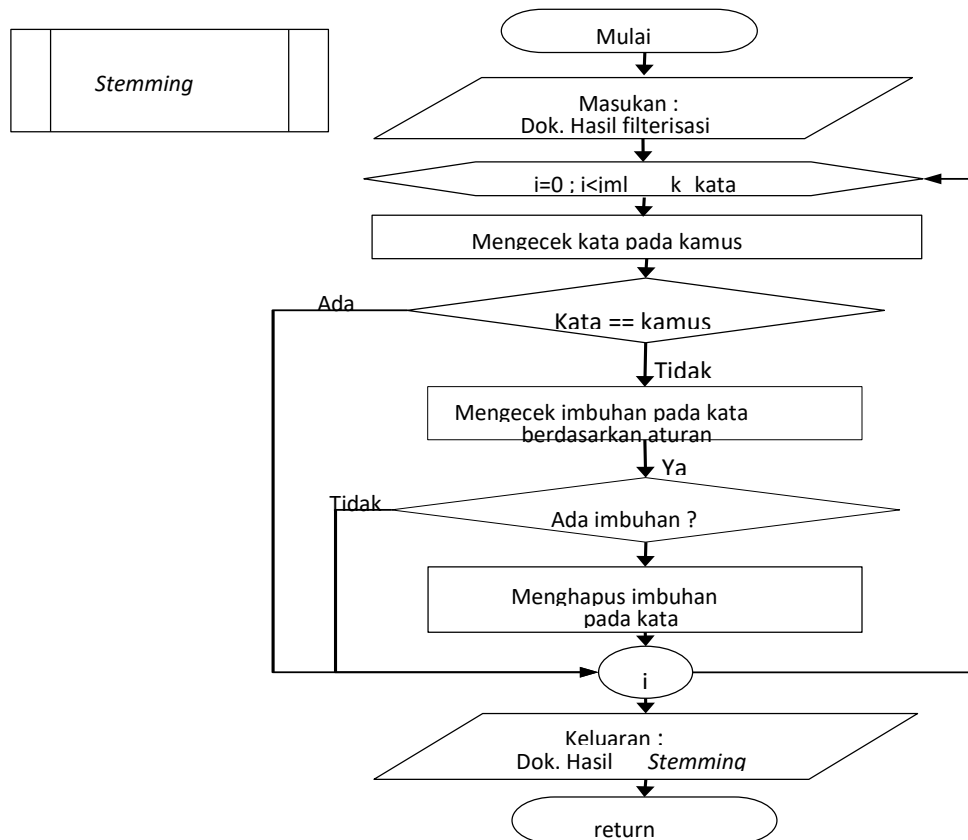
**Gambar 4.4 Diagram Alir Case Folding**



**Gambar 4.5 Diagram Alir Tokenizing**

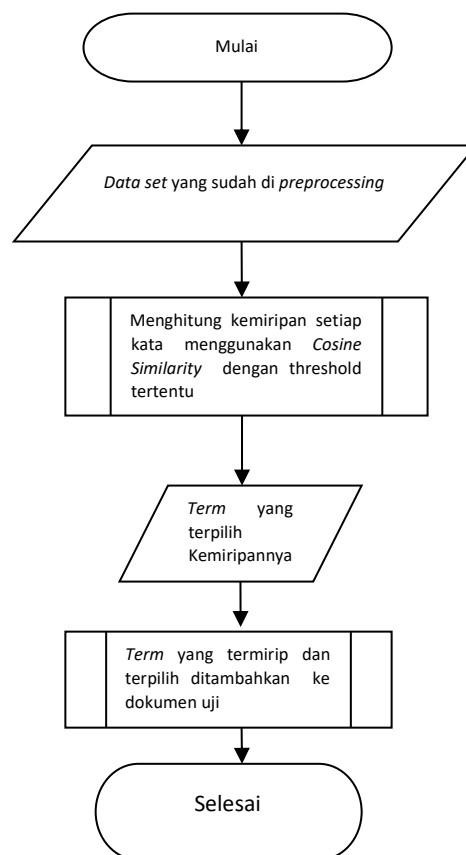


**Gambar 4.6 Diagram Alir *Filtering***



**Gambar 4.7 Diagram Alir *Stemming***

Proses yang dilakukan pada *preprocessing* terhadap data yang digunakan melalui beberapa tahapan, untuk menjadi kata yang dapat digunakan sebagai *term* dalam algoritme *query expansion* dan proses klasifikasi. *Preprocessing* memiliki empat tahapan yaitu *case folding*, *tokenizing*, *filtering*, serta *stemming*. Setelah melalui seleksi dari empat tahapan tersebut, selanjutnya akan dilakukan ekspansi kata. Untuk melakukan ekspansi kata dibutuhkan kamus berita sebagai dataset untuk memperoleh kedekatan kata atau *term* yang berhubungan dengan kata pada dokumen uji. Perhitungan yang digunakan yaitu menghitung jumlah frekuensi kemunculan kata, yang selanjutnya menghitung kemiripan *term* menggunakan teknik *Euclidean distance* untuk memperoleh *term* yang memiliki nilai terkecil untuk ditambahkan pada dokumen data uji. Proses pada algoritme tersebut digunakan, namun sebelumnya memperoleh *term* juga dilakukan proses *preprocessing* seperti pada Gambar 4.3. Hasil dari *query expansion* tersebut akan memperoleh kata yang terdekat dari kamus yang ditambahkan term-nya terhadap data uji, setelah itu akan diproses klasifikasi. Proses pada ekspansi kata dapat ditunjukkan oleh Gambar 4.8.

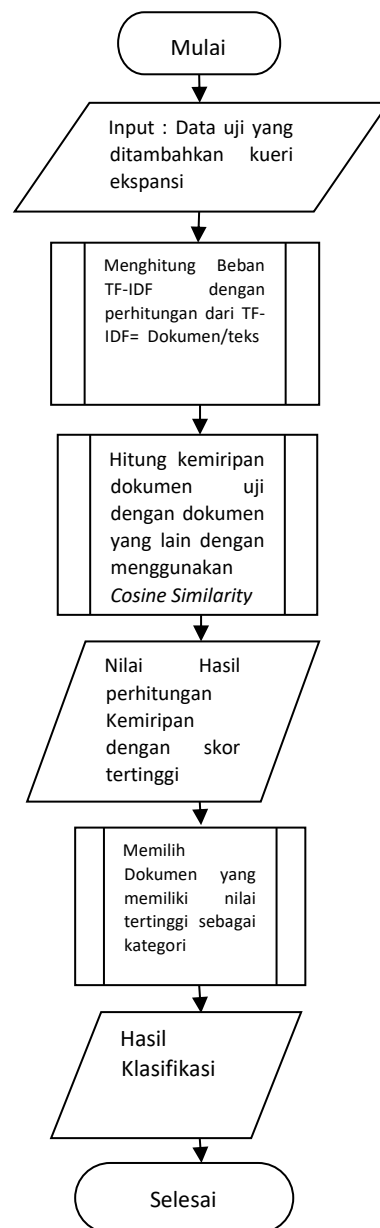


**Gambar 4.8 Diagram Alir Diagram alir Ekspansi Kata**

Proses ekspansi kata dilakukan setelah melalui proses *preprocessing* terlebih dahulu. *Preprocessing* dilakukan untuk menghasilkan kata yang selanjutnya akan digunakan pada proses pencarian kamus kedekatan kata dengan menghitung jumlah frekuensi kemunculan kata. Kemudian tahap perhitungan tersebut

menggunakan teknik *euclidean distance* untuk menghasilkan *term* yang memiliki nilai terkecil untuk dijadikan sebagai kueri ekspansi pada data uji dengan *term* yang terdapat pada kamus kedekatan kata tersebut. *Term* yang memiliki nilai tertinggi akan ditambahkan pada dokumen data uji tersebut.

Membuat kamus kedekatan kata diperlukan teknik untuk menentukan *term* yang terdekat untuk menghasilkan kata terdekat. Dari sekumpulan kata pada *term* di kamus berita dicari kata yang sama dari dokumen uji ,lalu menghitung jarak kata yang terdekat untuk dijadikan sebuah *query* yang akan menjadi ekspansi kata dan ditambahkan pada dokumen uji. Selanjutnya prosesnya adalah klasifikasi dokumen seperti pada Gambar 4.9.



**Gambar 4.9 Diagram Alir Klasifikasi K-Nearest Neighbor dengan *Query Expansion***

Klasifikasi teks dengan metode K-Nearest Neighbor dilakukan dengan beberapa tahapan seperti pada Gambar 4.9. Setelah melalui proses *preprocessing* dan ekspansi kata, selanjutnya akan menghitung *term* – dokumen dengan jumlah frekuensi kata dan menghitung bobot dengan teknik TF-IDF untuk menghasilkan nilai dari setiap kata yang terhubung dengan setiap dokumen. Kemudian untuk memperoleh hasil klasifikasi, diperlukan teknik untuk memperoleh ketentuan jenis dokumen. Teknik yang digunakan adalah menghitung jarak dokumen uji dengan dokumen latih yang memiliki nilai *Cosine Similarity*. Setelah memperoleh dokumen yang memiliki jarak terdekat, maka dapat disimpulkan bahwa dokumen tersebut memiliki jenis konten atau kategori yang sama, karena tingkat kemiripan kedua dokumen tersebut memiliki nilai paling maksimal.

### 4.3 Manualisasi Perhitungan Data

Manualisasi untuk menghitung dan menghasilkan nilai yang berguna untuk menjadi data untuk sistem. Perhitungan tersebut menggunakan dokumen kamus berita, dokumen data latih, serta dokumen uji untuk menghitung dan memproses klasifikasi.

#### 4.3.1 Manualisasi Klasifikasi Tanpa Ekspansi Kata

Manualisasi yang digunakan berikut merupakan perhitungan untuk menentukan klasifikasi dari sebuah *tweets* dari media sosial Twitter tanpa dilakukan ekspansi kata (*query expansion*). Proses tahap perhitungan klasifikasi tanpa menggunakan ekspansi kata menggunakan data latih pada Tabel 4.1.

**Tabel 4.1 Data Latih 1**

| Doc | Isi                                                              | Jenis Kategori |
|-----|------------------------------------------------------------------|----------------|
| 1   | Di Indonesia, Gangguan Tidur Banyak Dialami Pasien Sakit Jantung | Kesehatan      |
| 2   | Pantai Kucing, Primadona Wisata di Teluk Tomini                  | Travel         |
| 3   | Toyota: Masyarakat Mulai Beli Mobil Lagi                         | Otomotif       |
| 4   | Apple Perbaiki 100 Masalah iOS 9                                 | Teknologi      |
| 5   | Saus Keju Ini Praktis dan Enak Buat Cocolan Sayuran              | Makanan        |
| 6   | Forum Ekonomi Indonesia Dibentuk                                 | Ekonomi        |
| 7   | Potret Kehidupan John Lennon                                     | Entertainment  |
| 8   | Madrid Menang Tipis 2-1 atas Granada                             | Olahraga       |

Tabel 4.1 merupakan data uji yang berasal dari masing masing *tweets* dari setiap kategori yang diambil dari Twitter. Untuk Data Uji ditampilkan pada Tabel 4.2.

**Tabel 4.2 Data Uji Klasifikasi 2**

| Doc | Isi                                                    | Jenis Kategori    |
|-----|--------------------------------------------------------|-------------------|
| 9   | Oscar cetak gol pada menit ke-88. Shakhtar 2-1 Chelsea | Belum diketahui ? |

Proses yang dilakukan adalah langsung mengklasifikasi tanpa menggunakan fitur ekspansi kata pada dokumen *tweets* di atas, namun sebelumnya dokumen *tweets* tersebut telah dilakukan proses *preprocessing* terlebih dahulu. Berikut merupakan perhitungan klasifikasi yang menggunakan algoritme K-Nearest Neighbor pada Tabel 4.3.

**Tabel 4.3 Perhitungan *Tf-Idf* 1**

| Term       | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | df | Log(n/df) |
|------------|----|----|----|----|----|----|----|----|----|----|-----------|
| Indonesia  | 1  |    |    |    |    | 1  |    |    |    | 2  | 0,3010    |
| Ganggu     | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Banyak     | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Tidur      | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Dialami    | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Sakit      | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Pasien     | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Jantung    | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Pantai     | 0  | 1  |    |    |    |    |    |    |    | 1  | 0,9542    |
| Kucing     | 0  | 1  |    |    |    |    |    |    |    | 1  | 0,9542    |
| Primadona  | 0  | 1  |    |    |    |    |    |    |    | 1  | 0,9542    |
| Wisata     | 0  | 1  |    |    |    |    |    |    |    | 1  | 0,9542    |
| Teluk      | 0  | 1  |    |    |    |    |    |    |    | 1  | 0,9542    |
| Tomini     | 0  | 1  |    |    |    |    |    |    |    | 1  | 0,9542    |
| Toyota     | 0  |    | 1  |    |    |    |    |    |    | 1  | 0,9542    |
| Masyarakat | 0  |    | 1  |    |    |    |    |    |    | 1  | 0,9542    |
| Mulai      | 0  |    | 1  |    |    |    |    |    |    | 1  | 0,9542    |
| Beli       | 0  |    | 1  |    |    |    |    |    |    | 1  | 0,9542    |
| Mobil      | 0  |    | 1  |    |    |    |    |    |    | 1  | 0,9542    |

**Tabel 4.3 Perhitungan *Tf-Idf* 1 (lanjutan)**

|           |   |  |  |   |   |   |   |   |   |   |        |
|-----------|---|--|--|---|---|---|---|---|---|---|--------|
| Apple     | 0 |  |  | 1 |   |   |   |   |   | 1 | 0,9542 |
| Baik      | 0 |  |  | 1 |   |   |   |   |   | 1 | 0,9542 |
| 100       | 0 |  |  | 1 |   |   |   |   |   | 1 | 0,9542 |
| Masalah   | 0 |  |  | 1 |   |   |   |   |   | 1 | 0,9542 |
| iOs       | 0 |  |  | 1 |   |   |   |   |   | 1 | 0,9542 |
| 9         | 0 |  |  | 1 |   |   |   |   |   | 1 | 0,9542 |
| Saus      | 0 |  |  |   | 1 |   |   |   |   | 1 | 0,9542 |
| Keju      | 0 |  |  |   | 1 |   |   |   |   | 1 | 0,9542 |
| Praktis   | 0 |  |  |   | 1 |   |   |   |   | 1 | 0,9542 |
| Enak      | 0 |  |  |   | 1 |   |   |   |   | 1 | 0,9542 |
| Buat      | 0 |  |  |   | 1 |   |   |   |   | 1 | 0,9542 |
| Cocol     | 0 |  |  |   | 1 |   |   |   |   | 1 | 0,9542 |
| Sayur     | 0 |  |  |   | 1 |   |   |   |   | 1 | 0,9542 |
| forum     | 0 |  |  |   |   | 1 |   |   |   | 1 | 0,9542 |
| Ekonomi   |   |  |  |   |   | 1 |   |   |   | 1 | 0,9542 |
| Bentuk    |   |  |  |   |   | 1 |   |   |   | 1 | 0,9542 |
| Potret    |   |  |  |   |   |   | 1 |   |   | 1 | 0,9542 |
| Kehidupan |   |  |  |   |   |   | 1 |   |   | 1 | 0,9542 |
| John      |   |  |  |   |   |   | 1 |   |   | 1 | 0,9542 |
| Lennon    |   |  |  |   |   |   | 1 |   |   | 1 | 0,9542 |
| Madrid    |   |  |  |   |   |   |   | 1 |   | 1 | 0,9542 |
| Menang    |   |  |  |   |   |   |   | 1 |   | 1 | 0,9542 |
| Tipis     |   |  |  |   |   |   |   | 1 |   | 1 | 0,9542 |
| 2-1       |   |  |  |   |   |   |   | 1 |   | 1 | 0,9542 |
| Atas      |   |  |  |   |   |   |   | 1 |   | 1 | 0,9542 |
| Granada   |   |  |  |   |   |   |   | 1 |   | 1 | 0,9542 |
| oscar     |   |  |  |   |   |   |   |   | 1 | 1 | 0,9542 |
| Cetak     |   |  |  |   |   |   |   |   | 1 | 1 | 0,9542 |
| Gol       |   |  |  |   |   |   |   |   | 1 | 1 | 0,9542 |
| Menit     |   |  |  |   |   |   |   |   | 1 | 1 | 0,9542 |

**Tabel 4.3 Perhitungan *Tf-Idf* 1 (lanjutan)**

|          |  |  |  |  |  |  |  |  |   |   |        |
|----------|--|--|--|--|--|--|--|--|---|---|--------|
| 88       |  |  |  |  |  |  |  |  | 1 | 1 | 0,9542 |
| Shakhtar |  |  |  |  |  |  |  |  | 1 | 1 | 0,9542 |
| 2-1      |  |  |  |  |  |  |  |  | 1 | 1 | 0,9542 |
| Chelsea  |  |  |  |  |  |  |  |  | 1 | 1 | 0,9542 |

Perhitungan Tabel 4.3 merupakan untuk menghasilkan nilai *tf-df*, kemudian tahap selanjutnya adalah menghitung kemiripan antar dokumen data uji yang dicari kemiripannya dengan dokumen yang lain seperti pada Tabel 4.4.

**Tabel 4.4 Nilai Yang Digunakan Untuk menghitung Jarak Antar Dokumen 1.**

| Term       | D1     | D2     | D3     | D4     | D5     | D6     | D7 | D8 | D9 |
|------------|--------|--------|--------|--------|--------|--------|----|----|----|
| Indonesia  | 0,3010 |        |        |        |        | 0,3010 |    |    |    |
| Ganggu     | 0,9542 |        |        |        |        |        |    |    |    |
| Banyak     | 0,9542 |        |        |        |        |        |    |    |    |
| Tidur      | 0,9542 |        |        |        |        |        |    |    |    |
| Dialami    | 0,9542 |        |        |        |        |        |    |    |    |
| Sakit      | 0,9542 |        |        |        |        |        |    |    |    |
| Pasien     | 0,9542 |        |        |        |        |        |    |    |    |
| Jantung    | 0,9542 |        |        |        |        |        |    |    |    |
| Pantai     | 0,9542 | 0,9542 |        |        |        |        |    |    |    |
| Kucing     | 0      | 0,9542 |        |        |        |        |    |    |    |
| Primadona  | 0      | 0,9542 |        |        |        |        |    |    |    |
| Wisata     | 0      | 0,9542 |        |        |        |        |    |    |    |
| Teluk      | 0      | 0,9542 |        |        |        |        |    |    |    |
| Tomini     | 0      | 0,9542 |        |        |        |        |    |    |    |
| Toyota     | 0      |        | 0,9542 |        |        |        |    |    |    |
| Masyarakat | 0      |        | 0,9542 |        |        |        |    |    |    |
| Mulai      | 0      |        | 0,9542 |        |        |        |    |    |    |
| Beli       | 0      |        | 0,9542 |        |        |        |    |    |    |
| Mobil      | 0      |        | 0,9542 |        |        |        |    |    |    |
| Apple      | 0      |        |        | 0,9542 |        |        |    |    |    |
| Baik       | 0      |        |        | 0,9542 |        |        |    |    |    |
| 100        | 0      |        |        | 0,9542 |        |        |    |    |    |
| Masalah    | 0      |        |        | 0,9542 |        |        |    |    |    |
| iOs        | 0      |        |        | 0,9542 |        |        |    |    |    |
| 9          | 0      |        |        | 0,9542 |        |        |    |    |    |
| Saus       | 0      |        |        |        | 0,9542 |        |    |    |    |

**Tabel 4.4 Nilai Yang Digunakan Untuk menghitung Jarak Antar Dokumen 1.**

(lanjutan)

|              |   |  |  |  |        |        |        |        |        |
|--------------|---|--|--|--|--------|--------|--------|--------|--------|
| Keju         | 0 |  |  |  | 0,9542 |        |        |        |        |
| Praktis      | 0 |  |  |  | 0,9542 |        |        |        |        |
| Enak         | 0 |  |  |  | 0,9542 |        |        |        |        |
| Buat         | 0 |  |  |  | 0,9542 |        |        |        |        |
| Cocol        | 0 |  |  |  | 0,9542 |        |        |        |        |
| Sayur        | 0 |  |  |  | 0,9542 |        |        |        |        |
| forum        | 0 |  |  |  |        | 0,9542 |        |        |        |
| Ekonomi      |   |  |  |  |        | 0,9542 |        |        |        |
| Bentuk       |   |  |  |  |        | 0,9542 |        |        |        |
| Potret       |   |  |  |  |        |        | 0,9542 |        |        |
| Kehidupan    |   |  |  |  |        |        | 0,9542 |        |        |
| John         |   |  |  |  |        |        | 0,9542 |        |        |
| Lennon       |   |  |  |  |        |        | 0,9542 |        |        |
| Madrid       |   |  |  |  |        |        |        | 0,9542 |        |
| Menang       |   |  |  |  |        |        |        | 0,9542 |        |
| Tipis        |   |  |  |  |        |        |        | 0,9542 |        |
| 2-1          |   |  |  |  |        |        |        | 0,9542 |        |
| Atas         |   |  |  |  |        |        |        | 0,9542 |        |
| Granada      |   |  |  |  |        |        |        | 0,9542 |        |
| oscar        |   |  |  |  |        |        |        |        | 0,9542 |
| Cetak        |   |  |  |  |        |        |        |        | 0,9542 |
| Gol          |   |  |  |  |        |        |        |        | 0,9542 |
| Menit        |   |  |  |  |        |        |        |        | 0,9542 |
| 88           |   |  |  |  |        |        |        |        | 0,9542 |
| Shakhtar     |   |  |  |  |        |        |        |        | 0,9542 |
| 2-1          |   |  |  |  |        |        |        |        | 0,9542 |
| Chelsea      |   |  |  |  |        |        |        |        | 0,9542 |
| <b>Total</b> |   |  |  |  |        |        |        |        |        |

**Tabel 4.5 Normalisasi akhir 1**

| D1         | D2 | D3 | D4 | D5 | D6            | D7 | D8 | D9 |
|------------|----|----|----|----|---------------|----|----|----|
| 0.2352343  | 0  | 0  | 0  | 0  | 0.367560<br>6 | 0  | 0  | 0  |
| 0.34363222 | 0  | 0  | 0  | 0  | 0             | 0  | 0  | 0  |
| 0.34363222 | 0  | 0  | 0  | 0  | 0             | 0  | 0  | 0  |

**Tabel 4.5 Normalisasi akhir 1( lanjutan)**

|            |           |           |           |               |               |     |   |   |
|------------|-----------|-----------|-----------|---------------|---------------|-----|---|---|
| 0.34363222 | 0         | 0         | 0         | 0             | 0             | 0   | 0 | 0 |
| 0.34363222 | 0         | 0         | 0         | 0             | 0             | 0   | 0 | 0 |
| 0.34363222 | 0         | 0         | 0         | 0             | 0             | 0   | 0 | 0 |
| 0.34363222 | 0         | 0         | 0         | 0             | 0             | 0   | 0 | 0 |
| 0.34363222 | 0         | 0         | 0         | 0             | 0             | 0   | 0 | 0 |
| 0.34363222 | 0.4082483 | 0         | 0         | 0             | 0             | 0   | 0 | 0 |
| 0          | 0.4082483 | 0         | 0         | 0             | 0             | 0   | 0 | 0 |
| 0          | 0.4082483 | 0         | 0         | 0             | 0             | 0   | 0 | 0 |
| 0          | 0.4082483 | 0         | 0         | 0             | 0             | 0   | 0 | 0 |
| 0          | 0.4082483 | 0         | 0         | 0             | 0             | 0   | 0 | 0 |
| 0          | 0         | 0.4472136 | 0         | 0             | 0             | 0   | 0 | 0 |
| 0          | 0         | 0.4472136 | 0         | 0             | 0             | 0   | 0 | 0 |
| 0          | 0         | 0.4472136 | 0         | 0             | 0             | 0   | 0 | 0 |
| 0          | 0         | 0.4472136 | 0         | 0             | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0.4082483 | 0             | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0.4082483 | 0             | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0.4082483 | 0             | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0.4082483 | 0             | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0.4082483 | 0             | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0.4082483 | 0             | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0         | 0.37796<br>45 | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0         | 0.37796<br>45 | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0         | 0.37796<br>45 | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0         | 0.37796<br>45 | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0         | 0.37796<br>45 | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0         | 0.37796<br>45 | 0             | 0   | 0 | 0 |
| 0          | 0         | 0         | 0         | 0             | 0.536935<br>5 | 0   | 0 | 0 |
| 0          | 0         | 0         | 0         | 0             | 0.536935<br>5 | 0   | 0 | 0 |
| 0          | 0         | 0         | 0         | 0             | 0.536935<br>5 | 0   | 0 | 0 |
| 0          | 0         | 0         | 0         | 0             | 0             | 0.5 | 0 | 0 |

**Tabel 4.5 Normalisasi akhir 1 (lanjutan)**

|   |   |   |   |   |   |     |           |           |
|---|---|---|---|---|---|-----|-----------|-----------|
| 0 | 0 | 0 | 0 | 0 | 0 | 0.5 | 0         | 0         |
| 0 | 0 | 0 | 0 | 0 | 0 | 0.5 | 0         | 0         |
| 0 | 0 | 0 | 0 | 0 | 0 | 0.5 | 0         | 0         |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0.4082483 | 0         |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0.4082483 | 0         |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0.4082483 | 0         |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0.4082483 | 0         |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0.4082483 | 0         |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0.4082483 | 0         |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0         | 0.3535534 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0         | 0.3535534 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0         | 0.3535534 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0         | 0.3535534 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0         | 0.3535534 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0         | 0.3535534 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0         | 0.3535534 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0         | 0.3535534 |
|   |   |   |   |   |   |     |           |           |

Menghitung kemiripan antar dokumen normalisasi yang digunakan adalah dengan teknik *cosine similarity*, yaitu menghitung nilai dokumen data uji dari setiap *term* dengan dokumen yang lainnya untuk menghasilkan nilai tertinggi. Berikut merupakan hasil perhitungan menggunakan *cosine similarity* klasifikasi tanpa *query expansion*. Persamaan 4.1 merupakan penggunaan untuk menghitung *cosine similarity*.

$$Similarity = \cos(A, B) = \frac{A \cdot B}{||A|| ||B||} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (4.1)$$

| coSIM | D9 |
|-------|----|
| D1    | 0  |
| D2    | 0  |
| D3    | 0  |
| D4    | 0  |
| D5    | 0  |
| D6    | 0  |
| D7    | 0  |
| D8    | 0  |

Hasil yang diperoleh dari perhitungan manualisasi tanpa menggunakan *query expansion* didapatkan nilai yang tertinggi dengan D9 adalah belum dapat terklasifikasi karena tidak menemukan nilai yang tertinggi. Perhitungan tersebut dilakukan dan berlanjut sampai *term* terakhir. Untuk perhitungan data uji tersebut tidak hanya dilakukan dengan dokumen 1 (D1) saja, namun juga dilakukan dengan dokumen latih lain agar mendapatkan nilai dari masing masing perhitungan dan memperoleh nilai yang maksimal untuk diklasifikasikan sebagai kategori yang sama. Hal ini membuktikan bahwa dokumen uji pada D9 tidak mampu dikategorikan atau diklasifikasikan dengan baik karena masih belum memiliki tingkat akurasi yang bagus bahkan masih belum dapat diklasifikasikan. Permasalahan tersebut dapat diatasi agar mendapatkan akurasi yang lebih tinggi guna membantu proses klasifikasi, salah satunya adalah menggunakan metode *query expansion*.

#### 4.3.2 Manualisasi Menggunakan Query Expansion

Manualisasi untuk *query expansion* menggunakan kamus kedekatan kata yang berguna untuk memperoleh kata yang terdekat dari setiap kata pada data uji, dengan kamus tersebut dan layak dijadikan sebagai ekspansi kata. Namun sebelumnya, isi dari dokumen tersebut terlebih dahulu untuk diproses pada *preprocessing*. Contoh dokumen kamus yang digunakan ekspansi kata ditunjukkan pada Tabel 4.6.

**Tabel 4.6 Contoh Dokumen Kamus Berita**

| Doc | Isi                                                                     |
|-----|-------------------------------------------------------------------------|
| 1   | gol kemenangan, skor kini 1-1                                           |
| 2   | menang tipis dengan skor 2-0 atas chelsea, madrid ingin bungkam granada |
| 3   | madrid menang 2-0 atas granada                                          |
| 4   | madrid singkirkan granada                                               |
| 5   | granada tak sanggup kalahkan madrid                                     |
| 6   | beberapa pengusaha berencana membuat forum ekonom indonesia             |
| 7   | ekonom ingin usaha menengah keatas makin maju                           |
| 8   | android dan apple makin canggih                                         |
| 9   | teknologi ios7 makin benamkan os android                                |
| 10  | apple klaim android tiru ios7                                           |

Dari contoh yang diambil dari kamus pada Tabel 4.6 merupakan sebagian kecil dari seluruh dokumen kamus yang akan digunakan. Untuk lebih memudahkan pencarian data guna mengetahui kedekatan pada setiap kata,

maka diperlukan teknik untuk mengetahui kemunculan kata dan pertemuan setiap kata dengan kata yang lainnya secara bersamaan. Berikut merupakan *id* kata yang diekspansi pada Tabel 4.7. Sementara itu untuk Tabel 4.8 adalah menentukan jumlah kemunculan setiap kata pada setiap dokumen dan Tabel 4.9 merupakan jumlah kemunculan kata dan digunakan untuk menentukan kedekatan pada setiap kata.

**Tabel 4.7 *id* Kata Ekspansi**

| id | Kata      |
|----|-----------|
| 1  | menang    |
| 2  | skor      |
| 3  | chelsea   |
| 4  | madrid    |
| 5  | bungkam   |
| 6  | granada   |
| 7  | gol       |
| 8  | singkir   |
| 9  | sanggup   |
| 10 | kalah     |
| 11 | android   |
| 12 | apple     |
| 13 | canggih   |
| 14 | teknologi |
| 15 | ios7      |
| 16 | benam     |
| 17 | os        |
| 18 | klaim     |
| 19 | tiru      |
| 20 | ekonom    |
| 21 | usaha     |
| 22 | maju      |
| 23 | rencana   |
| 24 | buat      |
| 25 | forum     |

**Tabel 4.8 Jumlah Kata pada Setiap Dokumen**

|     | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|-----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1   | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 2   | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 3   | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 4   | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 5   | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 6   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 7   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 1  | 0  | 0  | 1  | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 8   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 1  | 1  | 0  | 0  | 1  | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  |
| 9   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0  |
| 10  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 0  | 1  | 1  | 1  |
| tot | 3 | 2 | 1 | 4 | 1 | 4 | 1 | 1 | 1 | 1  | 3  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 2  | 1  | 1  | 1  | 1  |

Tabel 4.8 menunjukkan kemunculan kata yang terdefinisi id(kolom) dengan setiap dokumen 1 hingga 10 (baris). Berikutnya adalah menghitung pertemuan setiap kata secara bersamaan seperti pada Tabel 4.9.

**Tabel 4.9 Perhitungan Kemunculan Antar Kata**

|    | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1  | 0 | 2 | 1 | 2 | 1 | 2 | 1 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 2  | 2 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 3  | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 4  | 2 | 1 | 1 | 0 | 1 | 4 | 1 | 1 | 1 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 5  | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 6  | 1 | 1 | 1 | 2 | 1 | 0 | 1 | 1 | 1 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 7  | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 8  | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 9  | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 10 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 0  | 2  | 1  | 1  | 2  | 1  | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  |

**Tabel 4.9 Perhitungan Kemunculan Antar Kata (lanjutan)**

|    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 1 | 1 |
| 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 1 | 1 |
| 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 |
| 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |

Setelah menghitung kemunculan kata, berikutnya adalah menghitung kedekatan jarak antar kedekatan kata. Misalnya menghitung id 7 dengan kata “gol” ,dan dihitung jaraknya dengan id 1 kata “menang”. Perhitungan tersebut menggunakan rumus *euclidean distance* seperti pada Persamaan 4.2.

$$d7, d1 = \sqrt{1-0)^2} + \sqrt{1-2)^2} + \sqrt{1-1)^2} + \dots \dots \quad (4.2)$$

Perhitungan tersebut dilakukan tidak hanya dengan d7 dengan d1, melainkan juga d7 dengan d yang lain untuk mencari jarak terpendek dengan nilai yang terkecil. Kemudian ekspansi dilakukan berdasarkan *threshold* tertentu yang berguna memberi nilai batas maksimal, dan memilih kata yang tidak melebihi nilainya dari *threshold* tersebut. Tujuan diberikannya *threshold* adalah agar kata yang tidak melebihi nilai dapat dijadikan ekspansi kata pada penambahan kata dalam dokumen yang akan diklasifikasikan. Tabel 4.10 menunjukkan hasil yang didapatkan dengan memperoleh kata ekspansi yang tidak melebihi *threshold* dari kamus kedekatan kata.

**Tabel 4.10 Hasil Ekspansi Kata**

| Doc | Kata    | Ekspansi Kata         |
|-----|---------|-----------------------|
| 1   | gol     | Menang<br>Skor madrid |
| 2   | chelsea | Madrid                |
| 3   | madrid  | Skor granada          |
| 4   | android | Apple ios             |
| 5   | ekonom  | usaha                 |

### 4.3.3 Manualisasi Klasifikasi Menggunakan *Query Expansion*

Manuliasasi pada perhitungan tanpa menggunakan ekspansi kata atau *query expansion* yang telah dilakukan pada sub bab sebelumnya menghasilkan hasil klasifikasi yang tidak sesuai dengan kategori aslinya. Selanjutnya pada sub bab ini akan menjelaskan manualisasi menggunakan ekspansi kata. Proses perhitungan klasifikasi dilakukan data latih pada Tabel 4.11.

**Tabel 4.11 Data Latih 2**

| Doc | Isi                                                              | Jenis Kategori |
|-----|------------------------------------------------------------------|----------------|
| 1   | Di Indonesia, Gangguan Tidur Banyak Dialami Pasien Sakit Jantung | Kesehatan      |
| 2   | Pantai Kucing, Primadona Wisata di Teluk Tomini                  | Travel         |
| 3   | Toyota: Masyarakat Mulai Beli Mobil Lagi                         | Otomotif       |
| 4   | Apple Perbaiki 100 Masalah iOS 9                                 | Teknologi      |
| 5   | Saus Keju Ini Praktis dan Enak Buat Cocolan Sayuran              | Makanan        |
| 6   | Forum Ekonomi Indonesia Dibentuk                                 | Ekonomi        |
| 7   | Potret Kehidupan John Lennon                                     | Entertainment  |
| 8   | Madrid Menang Tipis 2-1 atas Granada                             | Olahraga       |

Manualisasi ini menggunakan data latih, yang masing-masing kategori diambil salah satu contoh yang berasal dari *tweets*. Untuk melalui tahap proses klasifikasi, sebelumnya harus terlebih dahulu dilakukan ekspansi kata. Ekspansi kata dilakukan berdasarkan Tabel 4.12 berikut.

**Tabel 4.12 Daftar Ekspansi Kata**

| Doc | Kata    | Ekspansi Kata            |
|-----|---------|--------------------------|
| 1   | kulit   | dokter<br>sehat<br>cerah |
| 2   | gol     | Menang<br>skor           |
| 3   | chelsea | Madrid                   |

Data yang akan diuji untuk proses klasifikasi, sebelumnya telah dilakukan ekspansi kata pada Tabel 4.13. Setiap kata yang memiliki ekspansi kata langsung ditambahkan ke setelah kata yang bersangkutan.

**Tabel. 4.13 Data Uji Klasifikasi 2**

| Doc | Isi                                                                                                       | Keterangan            | Kategori |
|-----|-----------------------------------------------------------------------------------------------------------|-----------------------|----------|
| 1   | Oscar cetak gol pada menit ke-88. Shakhtar 2-1 Chlesea. #UCL                                              | Sebelum Ekspansi Kata | ?        |
| 2   | Oscar cetak gol <b>menang</b> <b>skor</b> pada menit ke-88. Shakhtar 2 <b>madrid</b> . #UCL<br>-1 Chlesea | Setelah Ekspansi Kata | ?        |

Manualisasi berikutnya adalah menghitung klasifikasi dokumen tersebut ,yaitu menghitung pembobotan dan penghitungan kedekatan jarak dengan dokumen latih berdasarkan Tabel 4.14.

**Tabel 4.14 Pembobotan TF-IDf 2**

| Term      | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | df | Log(n/df) |
|-----------|----|----|----|----|----|----|----|----|----|----|-----------|
| Indonesia | 1  |    |    |    |    | 1  |    |    |    | 2  | 0,6532    |
| Ganggu    | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Banyak    | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Tidur     | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Dialami   | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Sakit     | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Pasien    | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Jantung   | 1  |    |    |    |    |    |    |    |    | 1  | 0,9542    |
| Pantai    | 0  | 1  |    |    |    |    |    |    |    | 1  | 0,9542    |

**Tabel 4.14 Pembobotan TF-IDf 2 (lanjutan)**

|            |   |   |   |   |   |   |   |  |  |   |        |
|------------|---|---|---|---|---|---|---|--|--|---|--------|
| Kucing     | 0 | 1 |   |   |   |   |   |  |  | 1 | 0,9542 |
| Primadona  | 0 | 1 |   |   |   |   |   |  |  | 1 | 0,9542 |
| Wisata     | 0 | 1 |   |   |   |   |   |  |  | 1 | 0,9542 |
| Teluk      | 0 | 1 |   |   |   |   |   |  |  | 1 | 0,9542 |
| Tomini     | 0 | 1 |   |   |   |   |   |  |  | 1 | 0,9542 |
| Toyota     | 0 |   | 1 |   |   |   |   |  |  | 1 | 0,9542 |
| Masyarakat | 0 |   | 1 |   |   |   |   |  |  | 1 | 0,9542 |
| Mulai      | 0 |   | 1 |   |   |   |   |  |  | 1 | 0,9542 |
| Beli       | 0 |   | 1 |   |   |   |   |  |  | 1 | 0,9542 |
| Mobil      | 0 |   | 1 |   |   |   |   |  |  | 1 | 0,9542 |
| Apple      | 0 |   |   | 1 |   |   |   |  |  | 1 | 0,9542 |
| Baik       | 0 |   |   | 1 |   |   |   |  |  | 1 | 0,9542 |
| 100        | 0 |   |   | 1 |   |   |   |  |  | 1 | 0,9542 |
| Masalah    | 0 |   |   | 1 |   |   |   |  |  | 1 | 0,9542 |
| iOs        | 0 |   |   | 1 |   |   |   |  |  | 1 | 0,9542 |
| 9          | 0 |   |   | 1 |   |   |   |  |  | 1 | 0,9542 |
| Saus       | 0 |   |   |   | 1 |   |   |  |  | 1 | 0,9542 |
| Keju       | 0 |   |   |   | 1 |   |   |  |  | 1 | 0,9542 |
| Praktis    | 0 |   |   |   | 1 |   |   |  |  | 1 | 0,9542 |
| Enak       | 0 |   |   |   | 1 |   |   |  |  | 1 | 0,9542 |
| Buat       | 0 |   |   |   | 1 |   |   |  |  | 1 | 0,9542 |
| Cocol      | 0 |   |   |   | 1 |   |   |  |  | 1 | 0,9542 |
| Sayur      | 0 |   |   |   | 1 |   |   |  |  | 1 | 0,9542 |
| Forum      | 0 |   |   |   |   | 1 |   |  |  | 1 | 0,9542 |
| Ekonomi    |   |   |   |   |   | 1 |   |  |  | 1 | 0,9542 |
| Bentuk     |   |   |   |   |   | 1 |   |  |  | 1 | 0,9542 |
| Potret     |   |   |   |   |   |   | 1 |  |  | 1 | 0,9542 |
| Kehidupan  |   |   |   |   |   |   | 1 |  |  | 1 | 0,9542 |
| John       |   |   |   |   |   |   | 1 |  |  | 1 | 0,9542 |
| Lennon     |   |   |   |   |   |   | 1 |  |  | 1 | 0,9542 |

**Tabel 4.14 Pembobotan TF-IDf 2 (lanjutan)**

|          |  |  |  |  |  |  |  |   |   |   |        |
|----------|--|--|--|--|--|--|--|---|---|---|--------|
| Madrid   |  |  |  |  |  |  |  | 1 | 1 | 2 | 0,6532 |
| Menang   |  |  |  |  |  |  |  | 1 | 1 | 2 | 0,6532 |
| Tipis    |  |  |  |  |  |  |  | 1 |   | 1 | 0,9542 |
| 2-1      |  |  |  |  |  |  |  | 1 |   | 1 | 0,9542 |
| Atas     |  |  |  |  |  |  |  | 1 |   | 1 | 0,9542 |
| Granada  |  |  |  |  |  |  |  | 1 |   | 1 | 0,9542 |
| oscar    |  |  |  |  |  |  |  |   | 1 | 1 | 0,9542 |
| Cetak    |  |  |  |  |  |  |  |   | 1 | 1 | 0,9542 |
| Gol      |  |  |  |  |  |  |  |   | 1 | 1 | 0,9542 |
| Menit    |  |  |  |  |  |  |  |   | 1 | 1 | 0,9542 |
| 88       |  |  |  |  |  |  |  |   | 1 | 1 | 0,9542 |
| Shakhtar |  |  |  |  |  |  |  |   | 1 | 1 | 0,9542 |
| 2-1      |  |  |  |  |  |  |  |   | 1 | 1 | 0,9542 |
| Chelsea  |  |  |  |  |  |  |  |   | 1 | 1 | 0,9542 |
| Skor     |  |  |  |  |  |  |  |   | 1 | 1 | 0,9542 |

Setelah memperoleh hasil pembobotan dengan menggunakan *tf-idf*. Kemudian dokumen yang digunakan dalam perhitungan normalisasi berdasarkan Tabel 4.15.

**Tabel 4.15 Tabel Untuk Perhitungan Normalisasi 2**

| <i>Term</i> | D1     | D2     | D3 | D4 | D5 | D6     | D7 | D8 | D9 |
|-------------|--------|--------|----|----|----|--------|----|----|----|
| Indonesia   | 0,6532 |        |    |    |    | 0,6532 |    |    |    |
| Ganggu      | 0,9542 |        |    |    |    |        |    |    |    |
| Banyak      | 0,9542 |        |    |    |    |        |    |    |    |
| Tidur       | 0,9542 |        |    |    |    |        |    |    |    |
| Dialami     | 0,9542 |        |    |    |    |        |    |    |    |
| Sakit       | 0,9542 |        |    |    |    |        |    |    |    |
| Pasien      | 0,9542 |        |    |    |    |        |    |    |    |
| Jantung     | 0,9542 |        |    |    |    |        |    |    |    |
| Pantai      | 0,9542 | 0,9542 |    |    |    |        |    |    |    |
| Kucing      | 0      | 0,9542 |    |    |    |        |    |    |    |
| Primadona   | 0      | 0,9542 |    |    |    |        |    |    |    |
| Wisata      | 0      | 0,9542 |    |    |    |        |    |    |    |
| Teluk       | 0      | 0,9542 |    |    |    |        |    |    |    |

**Tabel 4.15 Tabel Untuk Perhitungan Normalisasi 2 (lanjutan)**

|            |   |        |        |        |        |        |        |        |        |
|------------|---|--------|--------|--------|--------|--------|--------|--------|--------|
| Tomini     | 0 | 0,9542 |        |        |        |        |        |        |        |
| Toyota     | 0 |        | 0,9542 |        |        |        |        |        |        |
| Masyarakat | 0 |        | 0,9542 |        |        |        |        |        |        |
| Mulai      | 0 |        | 0,9542 |        |        |        |        |        |        |
| Beli       | 0 |        | 0,9542 |        |        |        |        |        |        |
| Mobil      | 0 |        | 0,9542 |        |        |        |        |        |        |
| Apple      | 0 |        |        | 0,9542 |        |        |        |        |        |
| Baik       | 0 |        |        | 0,9542 |        |        |        |        |        |
| 100        | 0 |        |        | 0,9542 |        |        |        |        |        |
| Masalah    | 0 |        |        | 0,9542 |        |        |        |        |        |
| iOs        | 0 |        |        | 0,9542 |        |        |        |        |        |
| 9          | 0 |        |        | 0,9542 |        |        |        |        |        |
| Saus       | 0 |        |        |        | 0,9542 |        |        |        |        |
| Keju       | 0 |        |        |        | 0,9542 |        |        |        |        |
| Praktis    | 0 |        |        |        | 0,9542 |        |        |        |        |
| Enak       | 0 |        |        |        | 0,9542 |        |        |        |        |
| Buat       | 0 |        |        |        | 0,9542 |        |        |        |        |
| Cocol      | 0 |        |        |        | 0,9542 |        |        |        |        |
| Sayur      | 0 |        |        |        | 0,9542 |        |        |        |        |
| forum      | 0 |        |        |        |        | 0,9542 |        |        |        |
| Ekonomi    |   |        |        |        |        | 0,9542 |        |        |        |
| Bentuk     |   |        |        |        |        | 0,9542 |        |        |        |
| Potret     |   |        |        |        |        |        | 0,9542 |        |        |
| Kehidupan  |   |        |        |        |        |        | 0,9542 |        |        |
| John       |   |        |        |        |        |        | 0,9542 |        |        |
| Lennon     |   |        |        |        |        |        | 0,9542 |        |        |
| Madrid     |   |        |        |        |        |        |        | 0,9542 | 0,6532 |
| Menang     |   |        |        |        |        |        |        | 0,9542 | 0,6532 |
| Tipis      |   |        |        |        |        |        |        | 0,9542 |        |
| 2-1        |   |        |        |        |        |        |        | 0,9542 |        |
| Atas       |   |        |        |        |        |        |        | 0,9542 |        |
| Granada    |   |        |        |        |        |        |        | 0,9542 |        |
| oscar      |   |        |        |        |        |        |        |        | 0,9542 |
| Cetak      |   |        |        |        |        |        |        |        | 0,9542 |
| Gol        |   |        |        |        |        |        |        |        | 0,9542 |
| Menit      |   |        |        |        |        |        |        |        | 0,9542 |

**Tabel 4.15 Tabel Untuk Perhitungan Normalisasi 2 (lanjutan)**

|          |  |  |  |  |  |  |  |  |        |
|----------|--|--|--|--|--|--|--|--|--------|
| 88       |  |  |  |  |  |  |  |  | 0,9542 |
| Shakhtar |  |  |  |  |  |  |  |  | 0,9542 |
| 2-1      |  |  |  |  |  |  |  |  | 0,9542 |
| Chelsea  |  |  |  |  |  |  |  |  | 0,9542 |
| Skor     |  |  |  |  |  |  |  |  | 0,9542 |

**Tabel 4.16 Hasil Perhitungan Normalisasi 2**

| d1         | d2        | d3        | d4        | d5 | d6            | d7 | d8 | d9 |
|------------|-----------|-----------|-----------|----|---------------|----|----|----|
| 0.2352343  | 0         | 0         | 0         | 0  | 0.367560<br>6 | 0  | 0  | 0  |
| 0.34363222 | 0         | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0.34363222 | 0         | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0.34363222 | 0         | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0.34363222 | 0         | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0.34363222 | 0         | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0.34363222 | 0         | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0.34363222 | 0         | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0.34363222 | 0.4082483 | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0          | 0.4082483 | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0          | 0.4082483 | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0          | 0.4082483 | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0          | 0.4082483 | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0          | 0.4082483 | 0         | 0         | 0  | 0             | 0  | 0  | 0  |
| 0          | 0         | 0.4472136 | 0         | 0  | 0             | 0  | 0  | 0  |
| 0          | 0         | 0.4472136 | 0         | 0  | 0             | 0  | 0  | 0  |
| 0          | 0         | 0.4472136 | 0         | 0  | 0             | 0  | 0  | 0  |
| 0          | 0         | 0.4472136 | 0         | 0  | 0             | 0  | 0  | 0  |
| 0          | 0         | 0.4472136 | 0         | 0  | 0             | 0  | 0  | 0  |
| 0          | 0         | 0         | 0.4082483 | 0  | 0             | 0  | 0  | 0  |
| 0          | 0         | 0         | 0.4082483 | 0  | 0             | 0  | 0  | 0  |
| 0          | 0         | 0         | 0.4082483 | 0  | 0             | 0  | 0  | 0  |
| 0          | 0         | 0         | 0.4082483 | 0  | 0             | 0  | 0  | 0  |
| 0          | 0         | 0         | 0.4082483 | 0  | 0             | 0  | 0  | 0  |
| 0          | 0         | 0         | 0.4082483 | 0  | 0             | 0  | 0  | 0  |

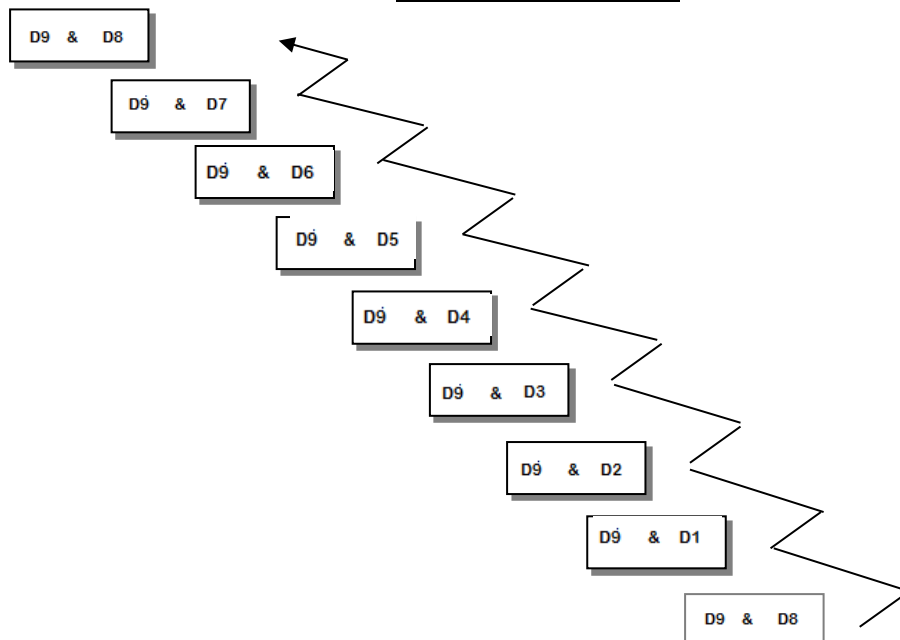
**Tabel 4.16 Hasil Perhitungan Normalisasi 2 (lanjutan)**

|   |   |   |   |               |               |     |               |           |
|---|---|---|---|---------------|---------------|-----|---------------|-----------|
| 0 | 0 | 0 | 0 | 0.37796<br>45 | 0             | 0   | 0             | 0         |
| 0 | 0 | 0 | 0 | 0.37796<br>45 | 0             | 0   | 0             | 0         |
| 0 | 0 | 0 | 0 | 0.37796<br>45 | 0             | 0   | 0             | 0         |
| 0 | 0 | 0 | 0 | 0.37796<br>45 | 0             | 0   | 0             | 0         |
| 0 | 0 | 0 | 0 | 0.37796<br>45 | 0             | 0   | 0             | 0         |
| 0 | 0 | 0 | 0 | 0.37796<br>45 | 0             | 0   | 0             | 0         |
| 0 | 0 | 0 | 0 | 0.37796<br>45 | 0             | 0   | 0             | 0         |
| 0 | 0 | 0 | 0 | 0             | 0.536935<br>5 | 0   | 0             | 0         |
| 0 | 0 | 0 | 0 | 0             | 0.536935<br>5 | 0   | 0             | 0         |
| 0 | 0 | 0 | 0 | 0             | 0.536935<br>5 | 0   | 0             | 0         |
| 0 | 0 | 0 | 0 | 0             | 0             | 0.5 | 0             | 0         |
| 0 | 0 | 0 | 0 | 0             | 0             | 0.5 | 0             | 0         |
| 0 | 0 | 0 | 0 | 0             | 0             | 0.5 | 0             | 0         |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0.408248<br>3 | 0.2171572 |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0.408248<br>3 | 0.2171572 |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0.408248<br>3 | 0         |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0.408248<br>3 | 0         |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0.408248<br>3 | 0         |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0.408248<br>3 | 0         |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0             | 0.317225  |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0             | 0.317225  |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0             | 0.317225  |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0             | 0.317225  |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0             | 0.317225  |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0             | 0.317225  |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0             | 0.317225  |
| 0 | 0 | 0 | 0 | 0             | 0             | 0   | 0             | 0.317225  |

Setelah dilakukan normalisasi pada Tabel 4.16, perhitungan selanjutnya yang dilakukan untuk menghasilkan nilai yang memiliki jarak terdekat yaitu D9 dengan

dokumen lainnya dan berikut adalah hasil perhitungan kedekatan jarak menggunakan KNN dengan teknik *cosine similarity*.

| coSIM | D9       |
|-------|----------|
| D1    | 0        |
| D2    | 0        |
| D3    | 0        |
| D4    | 0        |
| D5    | 0        |
| D6    | 0        |
| D7    | 0        |
| D8    | 0.177308 |



| No | 3 dokumen terdekat |
|----|--------------------|
| 1  | D9&D8 = 0,177308   |
| 2  | D9&D7 = 0          |
| 3  | D9&D6 = 0          |

Dengan perhitungan demikian, maka telah didapat hasil terdekat dengan D9 sebagai data uji, yaitu dengan D8 bernilai 0,177308 yang terkategori sama olahraga.

#### 4.4 Perancangan Antarmuka

Perancangan antarmuka pada penelitian ini merupakan tampilan yang akan dibangun nantinya sesuai perencanaan. Pada halaman yang akan dirancang ini

memiliki beberapa tampilan yaitu, tabel untuk memasukkan dokumen kamus berita yang digunakan untuk ekspansi kata dan memunculkan *term* uniknya, lalu terdapat juga halaman untuk klasifikasi teks dengan tabel data latih, data uji asli dan data uji setelah diproses, dan yang terakhir adalah halaman pengujian.

#### 4.4.1 Halaman Dokumen Kamus dan *Term* Unik

Pada halaman ini akan menampilkan dokumen yang dimasukkan sebagai dokumen kamus berita yang akan digunakan sebagai ekspansi kata. Proses ekspansi kata yang menghasilkan *term* unik ini telah melalui proses *preprocessing* oleh sistem terlebih dahulu. Berikut merupakan halaman yang akan direncanakan pada Gambar 4.10.

| DOKUMEN BERITA                                                                                                                                                                              |            | KLASIFIKASI |           | AKURASI |  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-------------|-----------|---------|--|
| <div style="display: flex; justify-content: space-between; align-items: center;"> <div style="border: 1px solid black; padding: 2px 5px;">INPUT</div> <div>UPLOAD KAMUS BERITA</div> </div> |            |             |           |         |  |
| id Dok                                                                                                                                                                                      | Isi Berita | id term     | term Unik |         |  |
|                                                                                                                                                                                             |            |             |           |         |  |

**Gambar 4.10 Halaman Perancangan Dokumen Kamus berita**

#### 4.4.2 Halaman Klasifikasi Teks

Halaman tampilan kedua akan menampilkan beberapa tabel yang terdapat data uji, data latih, dan data uji setelah diproses *preprocessing* dan ekspansi kata. Halaman ini bertujuan untuk memproses teks dengan klasifikasi setelah melalui tahap tersebut. Gambar 4.11 merupakan tampilan yang akan dirancang sesuai kebutuhan.

| DOKUMEN BERITA                    |                    | KLASIFIKASI | AKURASI     |
|-----------------------------------|--------------------|-------------|-------------|
| DOKUMEN LATIH                     |                    |             |             |
| ID L                              | ISI DOKUMEN TWEETS |             | KATEGORI ID |
|                                   |                    |             |             |
| DOKUMEN UJI (BELUM PREPROCESSING) |                    |             |             |
| ID UB                             | ISI DOKUMEN TWEETS |             |             |
|                                   |                    |             |             |
| DOKUMEN UJI (PREPROCESSING)       |                    |             |             |
| ID US                             | ISI DOKUMEN TWEETS |             |             |
|                                   |                    |             |             |

**Gambar 4.11 Perancangan Halaman Klasifikasi**

#### 4.4.3 Halaman Pengujian Klasifikasi

Perancangan pada halaman pengujian ini menampilkan beberapa tabel yang berisi hasil klasifikasi yang menggunakan ekspansi kata, dan juga ada hasil klasifikasi teks yang tanpa menggunakan ekspansi kata. Hal ini dilakukan untuk memberikan perbandingan hasil dari kedua proses tersebut. Pada Tabel 4.12 merupakan hasil dari klasifikasi yang dilakukan menggunakan *preprocessing* dan tanpa menggunakan *query expansion*, sedangkan tabel di bawahnya merupakan hasil yang diperoleh sistem dengan klasifikasi teks menggunakan *preprocessing* dan *query expansion*.

| DOKUMEN BERITA |                    | KLASIFIKASI   | AKURASI        |
|----------------|--------------------|---------------|----------------|
| ID             | ISI DOKUMEN TWEETS | KATEGORI ASLI | KATEGORI HASIL |
|                |                    |               |                |
| AKURASI        |                    |               |                |
|                |                    |               |                |
| ID             | ISI DOKUMEN TWEETS | KATEGORI ASLI | KATEGORI HASIL |
|                |                    |               |                |
| AKURASI        |                    |               |                |
|                |                    |               |                |

**Gambar 4.12 Perancangan Halaman Pengujian**

#### 4.5 Perancangan Pengujian

Pada pengujian yang akan dilakukan, akan dipilih 2 pengujian yaitu, pengujian akurasi maksimal nilai  $k$  dari K-Nearest Neighbor dan pengujian variasi *threshold* pada kedekatan ekspansi kata atau *query expansion* setelah diperolehnya nilai  $k$  tetangga terdekat dari metode K-Nearest Neighbor.

#### 4.5.1 Pengujian nilai $k$ pada KNN

Pengujian hasil akurasi nilai  $k$  tetangga terbaik dari K-Nearest Neighbor dilakukan dengan secara manual yang telah dilabeli dan berbentuk tabel untuk diketahui perbandingannya setiap nilai  $k$ . Tabel 4.17 merupakan perancangan pengujian akurasi nilai  $k$  pada KNN:

**Tabel 4.17 Perancangan Pengujian Nilai  $k$  Pada KNN**

| Nilai K | 1   | 3   | 5   | 7   | 9   | **  |
|---------|-----|-----|-----|-----|-----|-----|
| Akurasi | **% | **% | **% | **% | **% | **% |

#### 4.5.2 Perancangan Pengujian Variasi *Threshold* pada *Query Expansion*

Pengujian ini dilakukan setelah memperoleh nilai  $k$  terbaik dari pengujian sebelumnya. Pengujian ini dipilih untuk diketahui nilai batas atas atau *Threshold* ke berapa yang memiliki nilai akurasi tertinggi dan berhenti di nilai *Threshold* yang akurasinya mulai menurun. Tabel 4.18 merupakan perancangan pengujian.

**Tabel 4.18 Perancangan Pengujian Variasi *Threshold* pada *Query Expansion***

| Data Uji | Data Latih | Threshold yang digunakan | Akurasi Menggunakan Preprocessing dan <i>Query Expansion</i> |
|----------|------------|--------------------------|--------------------------------------------------------------|
| 100      | 400        | 0,0                      | **%                                                          |
| 100      | 400        | 0,1                      | **%                                                          |
| 100      | 400        | 0,2                      | **%                                                          |
| 100      | 400        | 0,3                      | **%                                                          |
| 100      | 400        | 0,4                      | **%                                                          |
| 100      | 400        | 0,5                      | **%                                                          |
| 100      | 400        | 0,6                      | **%                                                          |
| 100      | 400        | 0,7                      | **%                                                          |
| 100      | 400        | 0,8                      | **%                                                          |
| 100      | 400        | 0,**                     | **%                                                          |
| 100      | 400        | 0,**                     | **%                                                          |

#### **4.6 Penarikan Kesimpulan**

Penarikan kesimpulan dilakukan jika semua tahapan telah selesai dijalankan. Kesimpulan ini diambil dari rumusan masalah serta hasil pengujian yang dilakukan. Dilengkapi saran dan masukan guna pengembangan sistem pada peneliti berikutnya yang akan datang.