# BAB I

# PENDAHULUAN

## Latar Belakang

Perkembangan informasi digital pada era globalisasi saat ini mengalami peningkatan sehingga menyebabkan volume informasi berbentuk teks juga mengalami peningkatan. Informasi tersebut berasal dari berbagai aktivitas sumber yang terus meningkat, seperti *blog* maupun *website.* Informasi berbentuk teks ini bukan hanya sekitar 10.000 atau 100.000 informasi saja, melainkan bisa mencapai milyaran informasi. Kondisi ini akan menimbulkan terjadi melimpahnya informasi, sehingga menyebabkan pengguna informasi digital mengalami kesulitan dalam mencerna informasi yang ada, khususnya informasi teks dalam format tidak terstruktur. Selain itu, informasi teks dengan format tidak terstruktur juga menyebabkan lamanya proses pencarian informasi dengan topik tertentu.

Pelimpahan informasi berbentuk teks tidak terstruktur mendorong munculnya disiplin baru untuk menganalisis teks, yaitu *text mining. Text mining* adalah proses menganalisis teks dengan mencari pola-pola informasi yang digali dari suatu teks yang tidak terstruktur tersebut. Dari hasil menemukan pola teks tersebut dapat dilakukan proses klasifikasi untuk dokumen teks tersebut. Proses klasifikasi informasi teks dilakukan untuk mempermudah dalam proses pencarian dan mengelompokkan informasi sesuai dengan topiknya. Metode yang digunakan untuk proses klasifikasi sangatlah banyak, seperti metode *k-means,* *naïve bayes classifier, k-nearest neighbor, support vector machine (SVM), artificial neural network, fuzzy k-nearest neighbor,* dan lain sebagainya. Dari semua metode-metode yang ada, dipilihlah metode *naïve bayes classifier* dan *k-nearest neighbor* untuk membantu penelitian dalam mengklasifikasikan data-data tersebut. Pemilihan kedua algoritma ini dikarenakan algoritma ini memiliki nilai akurasi yang tinggi. Berdasarkan penelitian yang dilakukan oleh Hamzah didapat bahwa akurasi algoritma *naïve bayes classifier* sebesar 91% pada klasifikasi dokumen berita dan berdasarkan penelitian yang dilakukan oleh Samuel dkk didapatkan akurasi algoritma *k-nearest neighbor* sebesar 88,29% pada klasifikasi subtopik berita, selain itu kedua algoritma tersebut mudah untuk dipelajari.

*K-Nearest Neighbour* atau biasa disingkat dengan istilah KNN biasa digunakan dalam proses klasifikasi, salah satunya dalam klasifikasi teks. *K-Nearest Neighbour* merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain [1]. Untuk mengetahui jarak antar data baru dengan data lama dapat dilakukan dengan menghitung bobot data-data tersebut berdasarkan fitur yang ada, kemudian bobot tersebut akan dicocokan dan dihitung jaraknya.

*Naïve Bayes* merupakan suatu metode yang digunakan untuk mengklasifikasikan suatu objek berdasarkan perhitungan probabilitas. *Naïve Bayes* akan memprediksi suatu objek dengan memperhitungkan probabilitas suatu bukti untuk dijadikan sebagai hipotesis.

Penelitian kali ini akan membuat sistem yang dapat mengambil konten *website* berita *online* secara otomatis yang kemundian akan mengklasifikasikannya ke dalam topik tertentu. Untuk proses klasifikasi itu sendiri akan menggabungkan kedua algoritma diatas, yaitu *K-Nearest Neighbour* dan *Naïve Bayes*. Penelitian kali ini mengambil judul **“Implementasi *Naïve Bayes* dan *K-Nearest Neighbour* dalam Kategorisasi Artikel Berita di Internet”.** Pengambilan berita dari beberapa situs berita ini bertujuan untuk mengklasifikasikan berita-berita yang berasal dari beberapa situs berita sesuai dengan topik yang dibicarakan dalam artikel berita tersebut.

## Perumusan Masalah

Berdasarkan rumusan masalah yang dipaparkan di atas, maka diambilah suatu rumusan masalah sebagai berikut:

1. Implementasi metode *K-Nearest Neighbour* dan *Naïve Bayes* dalam mengkategorisasikan artikel dari situs berita *online*.
2. Mengukur akurasi yang didapat dari penggunaan metode *K-Nearest Neighbour* dan *Naïve Bayes* dalam mengkategorisasikan artikel dari situs berita *online*.

## Batasan Masalah

Penelitian ini memiliki batasan terhadap objek yang akan dibahas, mengingat adanya keterbatasan waktu dalam proses penelitian ini, selain itu juga agar penelitian yang akan dilakukan tidak melebar ke topik lainnya. Adapun batasan masalah dalam penelitian ini adalah:

1. Perangkat lunak yang akan dibuat berbasis *website* menggunakan Bahasa pemrograman PHPdan MySQL.
2. Artikel berita *online* yang digunakan merupakan artikel yang berasal dari situs berita [www.detik.com](http://www.detik.com), www.kompas.com, www.liputan6.com, www.viva.co.id, dan www. tribunnews.com.

## Tujuan Penelitian

Berdasarkan rumusan masalah yang dipaparkan di atas, maka tujuan dari penelitian ini adalah:

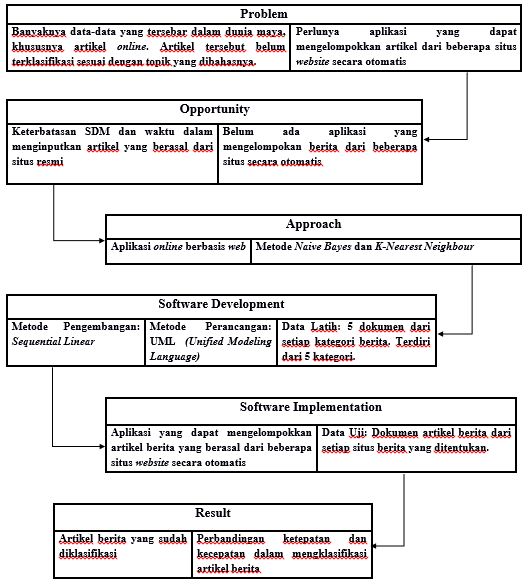
1. Untuk mengimplementasikan metode *K-Nearest Neighbour* dan *Naïve Bayes* dalam mengkategorisasikan artikel dari situs berita *online.*
2. Untuk mengetahui akurasi yang didapat dalam mengkategorikan artikel berita *online* dengan menggunakan metode *K-Nearest Neighbour* dan *Naïve Bayes.*

## Manfaat Penelitian

Manfaat dari penelitian ini diharapkan dapat mengimplementasikan suatu metode, yaitu metode *K-Nearest Neighbour* dan *Naïve Bayes* untuk mengkategorisasi artikel berita *online,* serta aplikasi yang dibuat dapat membantu dalam mengkategorikan artikel yang berasal dari situs berita *online* tanpa harus menggunakan proses manual. Selain itu juga dapat mengetahui nilai akurasi yang didapat dari penggunaan metode *K-Nearest Neighbour* dan *Naïve Bayes,* khususnya dalam mengkategorisasikan berita *online.*

## Kerangka Pemikiran

Adapun kerangka pemikiran dari Implementasi *Naïve Bayes* dan *K-Nearest Neighbour* dalam Kategorisasi Artikel Berita di Internet dapat dilihat pada Gambar 1.1 di bawah ini.

****

**Gambar 1.1** Kerangka Pemikiran

## Metode Penelitian

Metode penelitian ini terdiri dari dua tahap, yaitu tahap pengumpulan data dan tahap pengembangan perangkat lunak.

### Pengumpulan Data

Pada tahap pengumpulan data untuk membantu pembuatan aplikasi ini yaitu dengan melakukan beberapa cara, diantaranya:

1. Observasi

Observasi di sini dilakukan dengan meninjau langsung *website* berita *online*. Penijauan ini untuk mengamati data yang terdapat di *website* yang akan dijadikan sebagai bahan analisis kategorisasi.

1. Studi pustaka

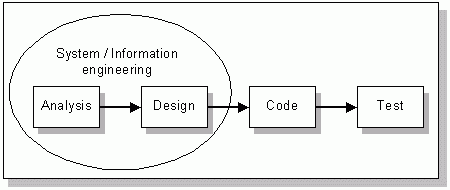
Studi pustaka di sini merupakan panduan dalam memecahkan masalah yang akan dibahas dengan mencari beberapa referensi sebagai landasan teoritis untuk menunjang penyusunan laporan tugas akhir ini, seperti buku cetak, jurnal, maupun referensi lainnya yang berasal dari internet.

1. Studi Dokumentasi

Studi dokumentasi dilakukan dengan mengamati dokumen-dokumen tertulis yang memiliki hubungan yang sama dalam menyelesaikan masalah.

### Pengembangan Perangakat Lunak

Model pengembangan perangkat lunak yang akan digunakan dalam membuat aplikasi ini yaitu model *sequential linear.* Model *sequential linear* biasa dikenal dengan model *waterfall* (air terjun) atau *classic life cycle* (alur hidup klasik). Model *sequential linear* menyediakan pendekatan alur hidup perangkat lunak secara terurut dimulai dari analisis, desain, pengkodean, pengujian, dan tahap pendukung *(support)* [7]. Tahap pengembangan perangkat lunak dengan model *sequential linear* dapat dilihat pada Gambar 1.2 di bawah ini.



**Gambar 1.2** Model *Sequential Linear*[[1]](#footnote-1)

1. Tahap analisis kebutuhan perangkat lunak adalah tahap untuk mengumpulkan kebutuhan perangkat lunak yang dilakukan dengan intensif. Tahap ini merupakan proses untuk menspesifikasikan kebutuhan perangkat lunak yang akan dibuat sehingga dapat dengan mudah dimengerti oleh *user.* Tahap ini perlu untuk didokumentasikan.
2. Tahap desain adalah tahap dimana kebutuhan perangkat lunak yang sebelumnya didapatkan pada tahap analisis akan digambarkan ke dalam desain perangkat lunak, seperti struktur data, arsitektur perangkat lunak, representasi antarmuka, dan prosedur pengkodean. Tahap ini perlu untuk didokumentasikan.
3. Tahap pengkodean adalah tahap menerjemahkan desain perangkat lunak yang dibuat pada tahap desain ke dalam kode-kode bahasa pemrograman. Hasil yang didapatkan dari tahap ini adalah program komputer yang sesuai dengan desain perangkat lunak yang sebelumnya dibuat pada tahap desain.
4. Tahap pengujian *(test)* adalah tahap untuk menguji fungsionalitas perangkat lunak untuk memastikan bahwa setiap bagian dari perangkat lunak tersebut berjalan sebagaiman mestinya. Tahap ini dilakukan untuk meminimalisir kesalahan *(error)*, serta untuk memastikan *output* yang dihasilkan sesuai dengan yang diinginkan.
5. Tahap pemeliharaan *(maintenance)* atau tahap pendukug *(support)* adalah tahap untuk pemeliharaan terhadap perangkat lunak yang sudah dikirimkan kepada *user* jika akan dilakukan perubahan atau perbaikan terhadap bagian perangkat lunak yang tidak berjalan sesuai keinginan *user*. Tahap ini dilakukan dengan mengulangi proses pengembangan yang dimulai dari tahap analisis untuk perubahan perangkat lunak yang sudah ada, tetapi tida untuk membuat perangkat lunak baru.

## Sistematika Penelitian

Sistematika penulisan laporan tugas akhir ini terdiri dari beberapa bab dengan masing-masing bab berisi uraian singkat dan penjelasan mengenai pokok pembahasan. Adapun sistematika penulisan laporan tugas akhir sebagai berikut:

**BAB I PENDAHULUAN**

Bab ini berisi mengenai latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, *state of the art,* kerangka pemikiran, metode penelitian, dan sistematika penelitian.

**BAB II LANDASAN TEORI**

Bab ini menjelaskan mengenai tinjauan pustaka dan penjelasan singkat dari beberapa teori-teori yang berhubungan dengan topik yang dibuat berdasarkan hasil penelitian dan hal-hal lainnya yang dapat membantu dalam proses penulisan laporan tugas akhir ini.

**BAB III ANALISIS DAN PERANCANGAN**

Bab ini berisi mengenai analisis dan perancangan aplikasi yang akan dibuat, yaitu berisi tentang identifikasi masalah, perancangan aplikasi yang terdiri dari desain sistem, basis data, dan antarmuka, cara kerja aplikasi, dan evalusi aplikasi.

**BAB IV IMPLEMENTASI DAN PENGUJIAN**

Bab ini berisi mengenai implementasi dari sistem yang dibangun baik itu yang berhubungan dengan *software* atau *hardware* yang dibutuhkan untuk mendukung aplikasi yang dibuat, selain itu juga dilakukan pengujian terhadap apliaksi yang dibuat.

**BAB V PENUTUP**

Bab ini berisi pembahasan singkat mengenai aplikasi yang dibuat secara keseluruhan serta saran untuk pengembangan perangkat lunak yang lebih baik untuk ke depannya.

# BAB II

# LANDASAN TEORI

Berdasarkan penjelasan pada BAB I bahwa isi dari BAB II akan membahas mengenai tinjauan pustaka dan uraian singkat dari beberapa teori-teori yang berhubungan dengan topik yang dibuat. Teori-teori tersebut berdasarkan pada penelitian atau hal-hal lainnya yang didapatkan dari buku, jurnal maupun dari sumber lainnya.

## Tinjauan Pustaka

Tinjauan pustaka berisi mengenai pemaparan penelitian terdahulu yang telah dilakukan oleh peneliti lain dan sudah dipublikasikan secara luas. Penelitian-penelitian yang pernah dilakukan mengenai *text mining* pada saat ini sudah banyak dilakukan oleh para peneliti di luar sana. Penelitian-penelitian tersebut menjelaskan bagaimana mengelola data berupa teks untuk dikelompokkan ke dalam beberapa kategori. Maka dari itu, penelitian yang akan dilakukan kali ini akan dibandingkan dengan beberapa *paper* yang membahas mengenai topik yang hampir sama dengan penelitian ini.Berikut adalah beberapa *paper* yang dijadikan sebagai referansi dalam melakukan penelitian untuk Tugas Akhir, diantaranya:

1. Andreas Daniel Arifin, Isye Arieshanti, dan Agus Zainal Arifin dengan *paper* yang berjudul “Implementasi Algoritma *K-Nearest Neighbour* yang Berdasarkan *One Pass Clustering* untuk Kategorisasi Teks”. *Paper* ini membahas mengenai teks-teks yang dikategorikan ke dalam kategori tertentu, serta membandingkan performa *one pass* KNN dengan KNN konvesional [2].
2. Yoseph Samuel, Rosa Delima, dan Antonius Rachmat dengan *paper* yang berjudul “Implementasi Metode *K-Nearest Neighbour* dengan *Decision Rule* untuk Klasifikasi Subtopik Berita”. *Paper* ini membahas mengenai berita-berita yang terdapat pada situs *website,* seperti bbc.com, cnn.com, dan foxnews.com akan dikategorikan berdasarkan topik olahraga yang terbagi menjadi tujuh subtopic [3].
3. Amir Hamzah dengan *paper* yang berjudul “Klasifikasi Teks dengan *Naïve Bayes Classifier (NBC)* untuk Pengelompokkan Teks Berita dan *Abstract* Akademis”. *Paper* ini meneliti performa algoritma NBC dalam kategorisasi teks yang berupa teks berita dan teks akademis berupa abstrak akademis dari berbagai disiplin ilmu [4].
4. Selvia Lorena Br Ginting dan Reggy Pasya Trinanda dengan *paper* yang berjudul “Penggunaan Metode *Naïve Bayes Classifier* pada Aplikasi Perpustakaan”. *Paper* ini membahas mengenai klasifikasi judul dan kategori buku di *database* perpustakaan agar memudahkan pengunjung untuk mengatahui *list* buku beserta tempat penyimpanannya [5].
5. Ahmad Hatta, Nana Ramadijanti, Afrida Helen dengan *paper* yang berjudul “Rancang Bangun Sistem Pengelolaan Dokumen-Dokumen Penting Menggunakan *Text Mining*”. *Paper* ini membahas mengenai pengelompokkan dokumen-dokumen penting, seperti KTP ke dalam tipe-tipenya. Dokumen tersebut akan diubah, yang sebelumnya dokumen berbentuk *hardcopy* (kertas) menjadi dokumen berbentuk data digital [6].

**Tabel 2.1** *State of The Art*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Peneliti** | **Studi Kasus** | **Metode Pemecahan Masalah** | ***Tools* yang digunakan** | **Rancangan Sistem** |
| Arifin, dkk | *Dataset Reuters-21578* | Algoritma *K-Nearest Neighbour* berdasarkan *One Pass Clustering* | *Java Runtime Environtment 1.6.0.29* dan *XAMPP 1.7.7* | Perancangan sistem berbasis objek |
| Samuel, dkk (2014) | Berita pada situs bbc.com, cnn.com, dan foxnews.com | *Metode K-Nearest Neighbour* dengan *Decision Rule* | *Visual Basic.NET 2010* | Perancangan sistem terstruktur |
| Hamzah (2012) | teks berita dan teks akademis berupa abstrak akademis dari berbagai disiplin ilmu | *Naïve Bayes Classifier (NBC)* | *Java* | Perancangan sistem berbasis objek |
| Ginting dan Trinanda | *List* buku di perpustakaan | Metode *Naïve Bayes Classifier* | *Java* | Perancangan sistem berbasis objek |
| Hatta, dkk | Dokumen-dokumen penting yang berbentuk *hardcopy* | Algoritma *K-Nearest Neighbour* | *Java* | Perancangan sistem berbasis objek |

**Tabel 2.1** *State of The Art* (Lanjutan)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Peneliti** | **Studi Kasus** | **Metode Pemecahan Masalah** | ***Tools* yang digunakan** | **Rancangan Sistem** |
| Siti Nurpadilah (2016) | Artikel berita dari detik.com, kompas.com, liputan6.com, viva.co.id, tribunnews.com | Algoritma *K-Nearest Neighbour* dan *Naïve Bayes Classifier* | PHP, MySQL | Perancangan sistem berbasis *website* |

## Landasan Teori

Teori-teori yang dijelaskan pada bab ini diantaranya yaitu artikel berita, internet, *website, scraping, codeigniter, text mining,* pembobotan tf-idf, *K-Nearest Neighbour, Naïve Bayes, website,* model pengembangan perangkat lunak *linear sequential, context diagram,* UML *(Unified Modeling Language), database,* dan *black box testing.*

### Artikel Berita

Artkel adalah karangan faktual mengenai suatu hal yang dibuat untuk dipublikasikan melalui media cetak seperti koran, majalah, buletin atapun melalui media elektronik seperti internet. Pembuatan artikel ini bertujuan untuk berbagi informasi kepada masyarakat. Artikel dibuat dengan melakukan penelitian terlebih dahulu terhadap objek yang akan dijadikan sebagai bahan pembahasan.

Berita adalah sebuah laporan yang berisi informasi mengenai suatu kejadian pada saat itu juga. Dalam Wikipedia Indonesia dijelaska bahwa definisi berita adalah informasi baru atau informasi mengenai sesuatu yang sedang terjadi, disajikan lewat bentuk cetak, siaran, internet, atau dari mulut ke mulut [8]. Tidak semua kejadian dapat dikatakan berita. Suatu kejadian dapat dikatakan sebagai berita jika kejadian itu merupakan kejadian yang luar biasa dan baru saja terjadi, selain itu kejadian tersebut juga memiliki dampak bagi masyarakat jika kejadian tersebut disebarluaskan.

Melalui kedua penjelasan diatas maka definisi dari artikel berita adalah suatu karangan atau laporan suatu kejadian tertentu, dimana kejadian tersebut merupakan kejadian yang luar biasa dan baru saja terjadi yang kemudian akan dipublikasikan baik melalui media cetak maupun media elektronik.

### Internet

Internet *(Interconnection Networking)* adalah sekumpulan komputer yang terhubung satu sama lain melalui protokol TCP/IP sehingga memungkinkan adanya pertukaran informasi dari komputer *server* ke komputer *client.* Komputer-komputer tersebut akan terhubung dan membentuk sebuah jaringan melalui jalur telekomunikasi seperti telepon, *wireless,* ataupun satelit.

TCP *(Transmission Control Protocol)* adalah protokol yang bertanggung jawab untuk mengirimkan data dari *host* sumber ke *host* tujuan dengan benar[9]*.* Selain itu TCP juga bertugas untuk mendeteksi kesalahan atau terjadi kehilangan data saat proses pengiriman data. Jika terjadi hal tersebut maka TCP akan melakukan pengiriman kembali data yang salah atau hilang itu ke *host* tujuan sampai data diterima dengan lengkap. IP *(Internet Protocol)* adalah protokol yang mengatur bagaimana suatu data dapat dikenal dan dikirim dari satu komputer ke komputer lain [9]. IP biasa digunakan untuk pengalamatan internet atau biasa dikenal dengan nama *IP address.*

1. Sejarah Internet

Sejarah terbentuknya internet dimulai pada tahun 1969. Pada saat itu, DARPA *(Defence Advance Research)* yang merupakan lembaga riset Departemen Pertahanan Amerika mendanai sebuah proyek untuk mengembangkan jaringan komunikasi data antar beberapa komputer. Tujuannya adalah untuk keperluan militer, yaitu untuk menghubungkan daerah-daerah penting agar bisa terhubung dalam menghadapi masalah jika terjadi serangan nuklir dan gangguan keamanan lainnya. Perkembangan jaringan tersebut menghasilkan ARPNet. Pada tahun 1972, organisasi ini mendemonstrasikan proyek ini di depan peserta *The First Internastional Conference on Computer Communication* dengan memasang 40 titik jaringan.

FTP *(File Transfer Protokol)* merupakan aplikasi internet pertama yang diciptakan, yang berfungsi untuk mentransfer atau mengirimkan file. Jumlah host di internet pada tahun 1984 lebih dari 1000 host dan pada tahun ini pula diperkenalkan DNS *(Domain Name System)* yang digunakan untuk mendefinisikan IP *(Internet Protocol)* agar lebih mudah untuk diingat.

Tahun 1986, U.S.NSF *(National Science Foundation)* yang merupakan Lembaga Ilmu Pengetahuan Nasional Amerika Serikat mendanai pembuatan jaringan TCP/IP yang didanai oleh organisasi NSFNet. Jaringan TCP/IP digunakan untuk menghubungkan universitas-universitas di Amerika Serikat dengan kecepatan transfer file sebesar 56 Kbps. Jaringan TCP/IP merupakan awal lahirnya perkembangan internet sampai saat ini.

Tahun 1987, lahirlah ISP *(Internet Service Provider)* yang diberi nama UUNet, yang merupakan provider utama internet. Tahun 1995, perkembangan internet semakin luas yang ditandai dengan adanya penyedia jasa internet (ISP). NSFNet yang sudah lama menjadi pusat internet kemudian diubah kembali menjadi jaringan untuk keperluan riset. Karena perubahan ini, maka lalu lintas data yang melalui Amerika dialihkan ke jaringan pusat ISP. Sementara itu NSPNet mengembangkan VBNS *(Very High Speed Backbone Network Service)*, yaitu jaringan dengan kecepatan sangat tinggi yaitu sebesar 622 Mbps (OC-12), yang menghubungkan lima komputer super. Seiring dengan berjalannya waktu, aplikasi internet semakin banyak dengan diciptakannya WAIS *(Wide Area Information Servers), Gopher,* dan WWW *(World Wide Web)* [10]*.*

### Website

*Website* adalah halaman situs berupa berkas teks *(plainteks)* yang diatur sedemikian rupa melalui instruksi berbasis HTML dan juga bahasa skrip lainnya yang saling berhubungan satu sama lain. Halaman *web* berisi informasi baik berupa teks, gambar, audio, maupun video. Halaman *web* tersimpan dalam *server web* dan terhubung secara global dengan komputer *client* melalui jaringan TCP/IP sebagai protokol pertukaran paket data.

*Website* terbentuk dari beberapa komponen yang membentuk *website* itu sendiri. Komponen-komponen pembentuk *website* diantaranya:

1. *HyperText Markup Language* (HTML)

*HyperText Markup Language* (HTML) adalah bahasa pemrograman yang digunakan untuk membuat halaman *web*. Dokumen HTML dapat dijalankan diberbagai sistem operasi. HTML merupakan bahasa pemrograman yang tidak *case sensitive* yang artinya tidak ada perbadaan antara huruf kecil dengan huruf kapital. HTML terdiri dari tag-tag yang berisi elemen pembangun halaman web, seperti untuk menebalkan huruf maka digunakan tag <b> atau <B>.

1. *Hypertext Preprocessor* (PHP)

*Hypertext Preprocessor* (PHP) merupakan bahasa pemrograman *server side* yang biasa digunakan oleh *developer* untuk membangun serta mengembangkan *website.* Untuk menjalankan php itu sendiri yaitu dengan menggunakan *web server.* PHP yang diciptakan oleh Rasmus Lerdorf pada tahun 1994 yang awalnya singkatan dari *Personal Home Page* yang kemudian berubah menjadi *Hypertext Preprocessor.* Versi terbaru PHP saat ini yaitu versi ke 5 atau yang sering kita sebut PHP5.

Untuk melakukan eksekusi program PHP dibutuhkan perangkat lunak *web server* yaitu Apache. Apache menggunakan protokol HTTP untuk melayani fasilitas *web/www.* Apache dapat dijalankan di berbagai sistem operasi seperti Windows, Linux, dan sistem operasi lainnya.

1. *Cascading Style Sheet* (CSS)

*Cascading Style Sheet* (CSS)adalah bahasa *style sheet* untuk mengatur dokumen HTML, seperti pengaturan *background,* margin, *font, border, teks*, gambar atau yang lainya, sehingga membuat tampilan web menjadi lebih indah. Menurut Juju, CSS secara sederhana adalah sebuah metode yang digunakan untuk mempersingkat penulisan tag HTML, seperti *font, color, text*, dan tabel menjadi lebih ringkas sehingga tidak terjadi pengulangan penulisan [10]. CSS bersifat lintas *platform* dimana *script* ini dapat dibaca oleh berbagai macam sistem operasi dan *browser*. CSS dikembangkan oleh W3C *(World Wide Web Consortium),* yaitu suatu organisasi yang mengembangkan teknologi internet pada tahun 1996 [11].

1. Javascript

Javascript adalah bahasa pemrograman web berbasis java. Kode javascript disisipkan ke dalam halaman *web* melalui tag-tag yang berisi *script.* Javascript bersifat *case sensitif,* artinya huruf kecil dengan huruf capital berbeda. Javascript juga bersifat *client side scripting* yang eksekusinya berada di sisi *client* yang berupa *browser* seperti Internet Explorer (IE), Mozilla Firefox, Google Chrome, Netscape, dan Opera [12].

1. Jquery

Jquery adalah *library* yang berisi kumpulan kode-kode Javascript yang berfungsi untuk mempermudah dan mempercepat dalam pembuatan kode Javascript. Fitur yang terdapat di jquery diantaranya seperti fitur untuk memanipulasi *class* atau *id* pada CSS, efek animasi, dan yang lainnya.

1. *My Structure Query Language* (MySQL)

*My Structure Query Language* (MySQL)adalah sebuah program pembuat dan pengelola *database* atau yang disebut dengan *DataBase Management System* (DBMS) [13]. DBMS ini bersifat *open source*, yang artinya para *developer* dapat memperolehnya secara gratis melalui alamat website <http://www.mysql.com>.

### Scraping

*Scraping* adalah teknik untuk mengambil isi dari halaman *website* dengan spesifik, seperti mengambil *link website* atau *link image.* Teknik *scraping* dapat dilakukan menggunakan kode-kode pada HTML DOM *(Document Object Model).* DOM adalah objek model standar untuk HTML dan XML yang bersifat *platform independent* [12]*.*

### *CodeIgniter (CI)*

*CodeIgniter* atau biasa dikenal dengan CI adalah *framework* atau kerangka kerja yang menggunakan model MVC *(model, view, controller)* untuk membangun sebuah *website* dinamis dengan menggunakan bahasa pemrograman PHP [14]. Di dalam codeigniter tersedia banyak *library* yang dapat digunakan hanya dengan memanggil *class library* yang sudah disediakan*.* Codeignitermerupakan *framework*  yang stabil dan cepat dalam proses eksekusi programnya karena menggunakan konsep MVC yang berjalan sesuai dengan tugasnya.

Konsep MVC merupakan alur kerja dari *framework codeigniter.* MVC adalah sebuah teknik pemrograman yang memisahkan antara alur, data dan antarmuka sistem. Penjelasan mengenai konsep MVC yaitu [14]:

1. ***Model***

*Model* merupakan bagaian yang berhubungan dengan pengelolaan *database,* seperti manipulasi data *(insert, update, delete, search),* menangani validasi pada bagian *controller.* *Model* tidak berhubungan langsung dengan bagian *view.*

1. ***View***

*View* merupakan bagaian yang menangani *presentation logic.* *View* biasanya berisi *file template* HTML. *View* berfungsi untuk menerima dan merepresentasikan data ke *user*. Bagian *view* diatur oleh *controller* dan tidak memiliki hubungan langsung dengan *model.*

1. ***Controller***

*Controller* merupakan bagian yang mengatur hubungan antara *model* dan *view*. *Controller* bertugas untuk menerima *request* dari *user* kemudian menentukan apa yang akan diproses oleh sistem.

### Text Mining

*Text mining* adalah proses menganalisis suatu dokumen berupa teks yang tidak terstruktur dalam jumlah besar untuk mendapatkan informasi. *Text mining* dilakukan dengan pencarian pola pada teks, yaitu dengan mencari kata-kata pada dokumen yang dapat dijadikan sebagai perwakilan dari satu dokumen kemudian akan dianalisis dengan dokumen lainnya dan dilihat keterhubungannya dengan dokumen yang lain [5]. Terdapat beberapa tahapan yang dilakukan dalam *text mining* yaitu sebagai berikut:

1. Pemecahan kalimat, yaitu proses memecah dokumen teks yang panjang menjadi kumpulan kalimat. Pemecahan kalimat dilakukan dengan dengan melihat pemisah kalimat seperti tanda titik (.), tanda tanya (?), dan tanda seru (!). Hasil pemecahan kalimat berupa kalimat-kalimat yang menyusun dokumen teks tersebut .
2. *Case folding,* adalah proses mengubah semua huruf pada teks menjadi huruf kecil serta menghilangkan semua karakter yang ada dalam teks selain huruf a-z, kecuali tanda pemecah kalimat seperti spasi, tab, dan *newline* (garis baru).
3. *Tokenizing,* adalah proses pemotongan kata-kata dalam teks yang menyusun kalimat/*string.* Pemotongan kata dilakukan dengan melihat pemisah kata seperti spasi, tab, atau *newline* (baris baru). Hasil dari *tokenizing* adalah kata-kata tunggal penyusun *string*/kalimat yang dimasukan.
4. *Filtering,* adalah proses penyaringan terhadap kata-kata hasil dari *tokenizing* dengan menghilangkan *stoplist.* *Stoplist* adalah kata yang tidak relevan atau tidak memiliki keterkaitan dengan tema, seperti kata “dari”, “di”, “yang”, dan lain-lain.
5. *Stemming,* adalah proses mencari akar *(root)* kata dari setiap kata penyusun dokumen yaitu dengan mengembalikan kata berimbuhan ke dalam bentuk kata dasar.
6. *Analizing,* adalah proses untuk menghitung bobot (w) dokumen untuk dilihat tingkat kemiripannya dengan dokumen lain dengan melihat *keyword* penyusun dokumen tersebut.

### Pembobotan *TF-IDF*

TF-IDF *(Term Frequency-Inverse Document Frequency)* adalah pembobotan suatu kata *(term)* dalam suatu dokumen agar dapat diproses lagi menggunakan algoritma lain. TF merupakan jumlah kemunculan *term* dalam sebuah dokumen, sedangkan IDF merupakan jumlah kemunculan *term* pada semua dokumen. Nilai IDF sebuah *term*  dihitung dengan menggunakan rumus:

(1)

D adalah jumlah dokumen sedangkan *dfi* adalah jumlah kemunculan *term* terhadap dokumen D.

Pembobotan terhadap masing-masing dokumen terhadap kata kunci *(query)* dapat dilakukan menggunakan persamaan:

(2)

Keterangan:

merupakan bobot dokumen ke-d terhadap term ke-t.

d merupakan dokumen ke-d.

t merupakan *term* ke-t dari kata kunci.

tf merupakan jumlah kemunculan kata dari dokumen d.

### K-Nearest Neighbour

*K-Nearest Neighbour* atau biasa dikenal dengan KNN adalah metode untuk mengklasifikasikan suatu objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Menurut Jumadi dan Winarko, *K-Nearest Neighbour* merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain [14].

Langkah-langkah yang dilakukan dalam *k-nearest neighbor* yaitu:

* + - 1. Inputkan data latih, data uji, dan nilai *k.*
      2. Hitung bobot untuk setiap term dari data latih dan data uji menggunakan rumus tf-idf.
      3. Hitung kemiripan setiap data latih yang sudah diklasifikasikan dengan data uji. Untuk menghitung kemiripan data latih dengan data uji menggunakan formula *cosine similarity,* yaitu:

(3)

Keterangan:

adalah *cosine similarity* dokumen d1 dan d2.

*t* adalah *term* atau kata ke-*t* dalam kalimat.

adalah bobot *term t* pada dokumen d1.

adalah bobot *term t* pada dokumen d2.

* + - 1. Urutkan hasil perhitungan kemiripan.
      2. Ambil data latih sebanyak *k* yang paling tinggi tingkat kemiripannya dengan data uji.
      3. Tentukan kelas berdasarkan kelas yang paling dominan dari *k* data latih yang sudah diambil.

Terdapat beberapa kondisi untuk menentukan kemiripan suatu data pada *k-nearest neighbor,* yaitu:

1. Jika terdapat 1 kelas dari 1 data latih sebagai tetangga terdekat (terdekat pertama) dari data uji, maka dijadikan sebagai kelas untuk data uji.
2. Jika dalam *k*-tetangga terdapat 2 kelas yang berbeda, maka diambil kelas dengan jumlah data terbanyak (mayoritas voting).
3. Jika terdapat 2 kelas dimana pertama memiliki jumlah tetangga jauh lebih banyak dan kelas kedua memiliki jumlah tetangga dekat lebih sedikit. Keadaan ini akan menyebabkan distorsi data. Untuk mengatasi mayoritas voting pada kasus ini maka ditambahkan penggunaan bobot untuk menghitung kandidat kelas yang sebaiknya diambil oleh data uji dari *k*-tetangga terdekat.

### Naïve Bayes

*Naïve bayes* adalah metode untuk mengklasifikasikan suatu objek yang menggunakan perhitungan probabilitas. Algoritma *Naïve bayes* dikemukakan oleh seorang ilmuwan asal Inggris, yaitu Thomas Bayes. Aturan dasar *Naïve bayes* adalah suatu hipotesis atau peristiwa dapat diperkirakan berdasarkan beberapa bukti yang diamati. Rumus umum *Naïve bayes* yaitu:

(6)

Keterangan:

P(H | E) : Probabilitas akhir bersyarat *(conditional probability)* suatu hipotesis H terjadi jika diberikan bukti *(evidence)* E terjadi.

P(E | H) : Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H.

P(H) : Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.

P(E) : Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain.

Rumus *Naïve bayes* untuk klasifikasi atau biasa disebut *Naïve Bayes Clasifier* (NBC) menggunakan formula di bawah ini:

(7)

Keterangan:

P(Y | X) : Probabilitas data dengan vektor X pada kelas Y.

P(Y) : Probabilitas awal kelas Y.

: Probabilitas independen kelas Y dari semua fitur dalam vektor X [1].

: Dibaca *phi*, merupakan simbol matematika dalam geometri euklidean yang berarti perbandingan (rasio) antara keliling lingkaran dengan diameternya. Dalam kasus ini untuk menghitung rasio dari titik ke-*i=1*dengan nilai *q* hingga titik ke*-i=n* dengan nilai *qn.*

Klasifikasi dengan metode *naïve bayes* dilakukan dengan menentukan nilai kategori yang paling tinggi probabilitasnya (VMAP) dari dokumen yang akan diklasifikasikan berdasarkan *term* yang muncul dalam dokumen. Penentuan nilai kategori dilakukan dengan memasukan atribut <a1, a2, a3, … an>. Dari atribut-atribut tersebut akan dicari nilai maksimumnya dengan rumus:

(8)

Berdasarkan teorema bayes pada persamaan (6), maka persamaan (8) ditulis:

(9)

untuk semua *vj* memiliki nilai yang sama, maka persamaan ini dapat ditulis:

(10)

Dengan mengasumsikan bahwa nilai setiap kata dalam atribut <a1, a2, a3, … an> sama, maka nilai dalam persamaan (10) ditulis:

(11)

Sehingga persamaan (10) ditulis:

(12)

dan probabilitas kata *wk* untuk setiap kategori dihitung pada saat pelatihan. Untuk mencari nilai dan dihitung menggunakan rumus:

(13)

(14)

Keterangan:

*|docs j|* merupakan banyaknya dokumen pada kategori j saat pelatihan.

*|Contoh|* merupakan banyaknya dokumen yang digunakan saat pelatihan.

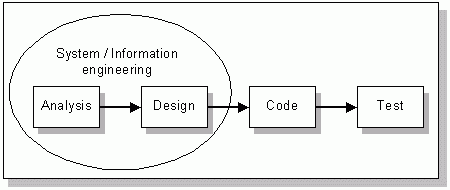
*nk* merupakan frekuensi kemunculan kata *wk* pada kategori *vj.*

*n* merupakan banyaknya kata pada semua dokumen dengan kategori *vj.*

*|kosakata|* merupakan banyaknya kata dalam pelatihan.

### Model Pengembangan Perangkat Lunak *Linear Sequential*

Model pengembangan perangkat lunak *linear sequential* atau biasa dikenal dengan model *waterfall* adalah model pengembangan perangkat lunak dengan pendekatan alur hidup perangkat lunak yang terurut *(sequential)*, dimulai dari analisis, desain, pengodean, pengujian, dan tahap pengujian.



**Gambar 2.1** Model *Sequential Linear*[[2]](#footnote-2)

1. Tahap analisis kebutuhan perangkat lunak adalah tahap untuk mengumpulkan semua kebutuhan perangkat lunak yang dilakukan secara intensif. Tahap ini merupakan proses untuk menspesifikasikan kebutuhan perangkat lunak yang akan dibuat sehingga dapat dengan mudah dimengerti oleh *user.* Tahap analisis perlu untuk didokumentasikan.
2. Tahap desain adalah tahap dimana kebutuhan perangkat lunak yang sebelumnya didapatkan pada tahap analisis akan digambarkan ke dalam desain perangkat lunak, seperti struktur data, arsitektur perangkat lunak, representasi antarmuka, dan prosedur pengkodean. Tahap ini perlu untuk didokumentasikan.
3. Tahap pengkodean adalah tahap menerjemahkan desain perangkat lunak yang dibuat pada tahap desain ke dalam kode-kode bahasa pemrograman. Hasil yang didapatkan dari tahap ini adalah program komputer yang sesuai dengan desain perangkat lunak yang sebelumnya dibuat pada tahap desain.
4. Tahap pengujian *(test)* adalah tahap untuk menguji fungsionalitas perangkat lunak untuk memastikan bahwa setiap bagian dari perangkat lunak tersebut berjalan sebagaiman mestinya. Tahap ini dilakukan untuk meminimalisir kesalahan *(error)*, serta untuk memastikan *output* yang dihasilkan sesuai dengan yang diinginkan.
5. Tahap pemeliharaan *(maintenance)* atau tahap pendukug *(support)* adalah tahap untuk pemeliharaan terhadap perangkat lunak yang sudah dikirimkan kepada *user* jika akan dilakukan perubahan atau perbaikan terhadap bagian perangkat lunak yang tidak berjalan sesuai keinginan *user*. Tahap ini dilakukan dengan mengulangi proses pengembangan yang dimulai dari tahap analisis untuk perubahan perangkat lunak yang sudah ada, tetapi tidak untuk membuat perangkat lunak baru.

### Unified Modeling Language (UML)

*Unified Modeling Language* (UML) adalah bahasa pemodelan standar pada pembangunan perangkat lunak berbasis objek untuk medefinisikan *requirement* perangkat lunak, membuat analisis dan desain perangkat lunak, serta menggambarkan arsitektur perangkat lunak. UML merupakan standar bahasa yang banyak digunakan untuk pengembangan perangkat lunak berorientasi objek.

UML dikeluarkan oleh perusahaan *Rational Software Corporation* pada tahun 1995. UML dibangun dari beberapa gabungan konsep pemodelan, seperti OMT *(Object Modelling Technique)* yang diciptakan oleh Rambaugh dan Booch (1991), CRC *(The Classes, Responsibilities, Collaborators)* oleh Rebecca Wirfs Brock (1990), dan konsep lainnya [7]. Dari gabungan konsep-konsep tersebut, UML mengambil elemen-elemen yang lebih efektif dan membuang elemen-elemen yang tidak praktis, serta menambahkan elemen-elemen baru yang sebelumnya belum ada pada konsep terdahulu sehingga UML lebih ekspresif dari konsep pemodelan yang lainnya.

1. Use Case Diagram

*Use case diagram* adalah pemodelan pada sistem berorientasi objek yang menggambarkan kelakuan *(behaviour)* dari sistem yang akan dibuat. *Use case* mendeskripsikan interaksi antara aktor dengan sistem. Dengan kata lain, *use case* digunakan untuk mengetahui siapa saja yang berhak menggunakan sistem dan fungsi apa saja yang dapat dilakukan oleh sistem.

Terdapat aturan dalam pemberian nama pada *use case,* yaitu *use case* harus didefinisikan sesimpel mungkin dan dapat dipahami oleh para *developer*. Aturan untuk pemberian nama pada aktor harus menggunakan kata benda sedangkan pemberian nama pada *use case* harus menggunakan kata kerja.

*Use case diagram* memiliki beberapa simbol yang digunakan untuk menggambarkan fungsionalitas sistem. Table 2.1 di bawah ini akan menjelaskan mengenai simbol-simbol yang terdapat pada *use case diagram.*

**Tabel 2.1** Simbol *Use Case Diagram*

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Nama** | **Simbol** | **Keterangan** |
| 1 | *Use case* | Description: Description: Description: Description: Description: Description: Description: Description: 2010-06-01_110252 | *Use case* mendefinisikan fungsionalitas yang disediakan oleh sistem. Pemberian nama pada *use case* menggunakan kata kerja di awal frase namanya. |
| 2 | Aktor/*actor* | Description: Description: Description: Description: Description: Description: Description: Description: 2010-06-01_110124 | *Actor* mendefinisikan orang, proses atau sistem lain yang berinteraksi dengan sistem yang akan dibuat. Pemberian nama pada *actor* menggunakan kata benda di awal frase namanya. |
| 3 | Asosiasi */ Association* | Description: Description: Description: Description: Description: Description: Description: Description: 2010-06-01_110406 | Asosiasi mendefinisikan interaksi antara *actor* yang berdengan *use case.* Asosiasi digambarkan dengan sebuah garis lurus yang menghubungkan *actor* dengan *use case.* |
| 4 | *Dependency include* |  | *Dependency include* mendefinisikan hubungan antara dua *use case* dimana *use case* yang satu akan memanggil *use case* yang lain. |
| 5 | Generalisasi |  | Generalisasi mendefinisikan hubungan antara dua *use case* dimana fungsi yang satu merupakan fungsi yang lebih umu dari fungsi yang lainnya. |

1. Class Diagram

*Class diagram* adalah pemodelan pada sistem berorientasi objek yang yang menggambarkan struktur sistem yang didefinisikan oleh kelas-kelas yang akan dibuat untuk membangun sistem tersebut. Kelas-kelas pada struktur sistem harus dapat melakukan fungsi sesuai dengan kebutuhan sistem.

*Class diagram* menggambarkan struktur kelas suatu sistem. Susunan struktur kelas yang baik pada *class diagram* memiliki jenis kelas sebagai beritkut:

1. Kelas *main*, merupakan kelas yang paling pertama dieksekusi ketika program dijalankan.
2. Kelas *view*, merupakan kelas yang mengatur tampilan *user interface.*
3. Kelas *controller,* merupakan kelas yang menangani fungsi-fungsi yang sudah didefinisikan sebelumnya pada *use case diagram.*
4. Kelas *model*, merupakan kelas yang menghubungkan sistem dengan *database* sehingga sistem dapat menambahkan, mengubah, atau menghapus data di *database.*

Tabel 2.2 di bawah ini merupakan simbol-simbol yang terdapat pada *class diagram.*

**Tabel 2.2** Simbol *Class Diagram*

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Nama** | **Simbol** | **Keterangan** |
| 1 | *Class* | Description: Description: Description: Description: Description: Description: Description: Description: Class | *Class* menggambarkan struktur perangkat lunak yang berisi kode programyang terkait dengan nama kelas. Class terdiri dari *object name* sebagai nama kelas, *attribute* sebagai variabel-variabel yang terdapat pada kelas, dan *operation* sebagai metode atau fungsi yang terdapat pada kelas. |
| 2 | *Association* | Description: Description: Description: Description: Description: Description: Description: Description: 2010-06-01_104903 | *Association* menggambarkan relasi antara dua kelas. *Association* digambarkan dengan garis lurus yang menghubungkan antara kelas yang satu dengan kelas lainnya. |
| 3 | *Composition* | Description: Description: Description: Description: Description: Description: Description: Description: 2010-06-01_105039 | *Composition* menggambarkan relasi antara satu kelas dengan kelas lain yang merupakan bagian dari kelas tersebut, |

**Tabel 2.2** Simbol *Class Diagram* (Lanjutan)

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Nama** | **Simbol** | **Keterangan** |
|  |  |  | dalam artian kelas pertama bergantung pada kelas kedua. |
| 4 | *Dependency* | Description: Description: Description: Description: Description: Description: Description: Description: 2010-06-01_105127 | *Dependency* menggambarkan operasi pada satu kelas yang menggunakan kelas lain. |
| 5 | *Aggregation* | Description: Description: Description: Description: Description: Description: Description: Description: 2010-06-01_105208 | *Aggregation* menggambarkan relasi antar kelas dengan makna “semua-bagian” |
| 6 | *Generalization* | Description: Description: Description: Description: Description: Description: Description: Description: 2010-06-01_105257 | *Generallization* menggambarkan relasi antar kelas dengan makna umum-khusus. |

1. Activity Diagram

*Activity diagram* adalah pemodelan pada sistem berorientasi objek yang menggambarkan alur kerja *(workflow)* atau aktivitas suatu proses pada perangkat lunak. Yang perlu diperhatikan disini adalah bahwa aktivitas yang dilakukan bukanlah aktivitas yang dilakukan aktor terhadapat sistem, tetapi aktivitas yang dapat dilakukan oleh sistem.

Pemodelan pada *activity diagram* diawali dengan *start* dan diakhiri dengan *end.* Pemodelan itu dibuat dengan simbol-simbol *activity diagram.* Table 2.3 di bawah ini menjelaskan simbol-simbol yang ada pada *activity diagram.*

**Tabel 2.3** Simbol *Activity Diagram*

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Nama** | **Simbol** | **Keterangan** |
| 1 | Status awal |  | Status awal aktivitas dari sistem. |
| 2 | Aktivitas |  | Aktivitas/alur kerja yang dilakukan oleh sistem. Penamaan aktivitas diawali oleh kata kerja. |

**Tabel 2.3** Simbol *Activity Diagram* (Lanjutan)

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Nama** | **Simbol** | **Keterangan** |
| 3 | Percabangan / *decision* |  | Asosiasi percabangan digambarkan jika terdapat pilihan aktivitas yang lebih dari satu. |
| 4 | Status akhir |  | Status akhir aktivitas yang dilakukan sistem. |
| 5 | *swimlane* | Description: Description: Description: Description: Description: Description: Description: Description: I:\SI Enterprise\2010-05-22_114414.png | *Swimlane* menggambarkan pemisahan pada organisasi bisnis yang bertanggung jawab pada aktivitas yang terjadi. |

1. Sequence Diagram

*Sequence diagram* adalah pemodelan pada sistem berorientasi objek yang menggambarkan kelakuan *(behaviour)* objek pada *use case* yang diddeskripsikan dengan waktu hidup objek serta *message* yang dikirimkan dan diterima dari objek satu ke objek lainnya. Untuk pembuatan *sequence diagram* maka harus mengetahui objek apa saja yang terlibat pada *use case* serta metode apa saja yang dimiliki oleh kelas yang diinstansiasikan menjadi objek tersebut. Selain itu, dalam pembuatan *sequence diagram* juga perlu melihat scenario yang ada pada *use case.*

Pemodelan *sequence diagram* juga dilakukan menggunakan simbol-simbol yang terdapat pada *sequence diagram.* Table 2.4 di bawah ini akan menjelaskan mengani simbol-simbol yang terdapat pada *sequence diagram.*

**Tabel 2.4** Simbol *Sequence Diagram*

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Nama** | **Simbol** | **Keterangan** |
| 1 | Aktor | Description: Description: Description: Description: Description: Description: Description: Description: 2010-06-01_110124  atau | Orang, proses, atau sistem lain yang berinteraksi dengan sistem yang akan dibangun. Penamaannya menggunakan kata benda di awal frase. |
| 2 | *Lifeline* |  | *Lifeline* menggambarkan kehidupan objek. |
| 3 | Waktu aktif |  | Menyatakan objek dalam keadaan aktif dan berinteraksi dengan sistem. Semua yang terhubung dengan waktu aktif menyatakan tahapan yang dilakukan di dalam sistem. |
| 4 | *Message* |  | *Messasge* menyatakan data atau proses yang dikirimkan atau diterima antar objek. |

1. Statechart Diagram

*Statechart diagram* adalah pemodelan pada sistem berorientasi objek yang menggambarkan daur hidup sistem dari awal sistem diinisialisasikan hingga keadaan *destroy. Statechart diagram* menggambarkan perubahan atau transisi suatu sistem atau objek. Jika *sequence diagram* digunakan untuk untuk interaksi antar objek, maka *statechart diagram* digunakan untuk interaksi di dalam sebuah objek [7].

*Statechart* diagram juga memiliki simbol-simbol yang digunakan untuk melakukan pemodelan pada sistem yang akan dibuat. Tabel 2.5 di bawah ini menjelaskan mengenai simbol-simbol pada *statecharrt diagram.*

**Tabel 2.5** Simbol *Statechart Diagram*

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Nama** | **Simbol** | **Keterangan** |
| 1 | *Start /* status awal |  | *Start* menggambarkan keadaan awal sistem mulai hidup. |
| 2 | *End /* status akhir |  | *End* menggambarkan keadaan akhir dari alur hidup sistem. |
| 3 | *Event* | Description: Description: Description: Description: Description: Description: Description: Description: C:\Users\aminus\Pictures\event.png | *Event* menggambarkan keadaan yang menyebabkan perubahan status mesin/objek. |
| 4 | *State* |  | *State* menggambarkan keadaan mesin atau objek pada saat tertentu. |

### Database

*Database* atau basis data adalah suatu kumpulan data yang saling terhubung *(interlated data)* dan disimpan secara bersama-sama pada suatu media [15]. *Database* terdiri dari kumpulan tabel-tabel yang saling berelasi satu sama lain. Tujuan dibentuknya tabel-tabel di sini adalah untuk menyimpan data ke dalam tabel-tabel agar mudah untuk diakses.

Menurut Sutanta dalam bukunya yang berjudul “Sistem Basis Data” menjelaskan bahwa suatu basis data memiliki beberapa kriteria penting yang harus dipenuhi, diantaranya:

Berorientasi pada data dan bukan berorientasi pada program yang akan digunakan. Artinya *database* harus disimpan secara terpisah dengan program aplikasinya.

Data dapat digunakan oleh *user* yang berbeda-beda atau beberapa program aplikasi tanpa harus mengubah *database.*

Data dalam *database* dapat berkembang dengan mudah baik volume maupun strukturnya.

Data yang ada dapat memenuhi kebutuhan sistem-sistem baru secara mudah.

Data dapat digunakan dengan cara yang berbeda-beda.

Kerangkapan data *(data redudancy)* minimal. Artinya data dalam *database* tidak perlu disimpan secara berulang. Hal ini dilakukan untuk menghindari kesulitan dalam pengolahan data di kemudian hari.

### Black Box Testing

*Black box testing* adalah pengujian yang fokus pada kebutuhan fungsionalitas sistem yang dibuat. *Black box testing* yaitu menguji perangkat lunak dari segi spesifikasi fungsional tanpa menguji desain dan kode program [7]. Pengujian dilakukan dengan maksud untuk mengetahui apakah fungsi-fungsi yang terdapat pada sistem sudah sesuai dengan spesifikasi yang dibutuhkan atau belum.

*Black box testing* dilakukan dengan membuat studi uji untuk dicoba semua fungsinya dengan memakai sistem tersebut apakah sesuai dengan spesifikasi yang dibutuhkan atau tidak. Studi uji yang dibuat terdiri dari kasus benar dan kasus salah. Contohnya seperti proses *login,* yaitu:

1. Jika *user* menggunakan *username* dan *password* yang benar.
2. Jika *user* menggunakan *username* dan/atau *password* yang salah.

Kategori yang diujikan dalam *black box testing* yaitu:

1. Fungsi-fungsi yang tidak benar atau hilang.
2. Kesalahan interface.
3. Kesalahan dalam struktur data atau akses *database* eksternal.
4. Kesalahan kinerja.
5. Inisialisasi dan kesalahan terminasi [16].

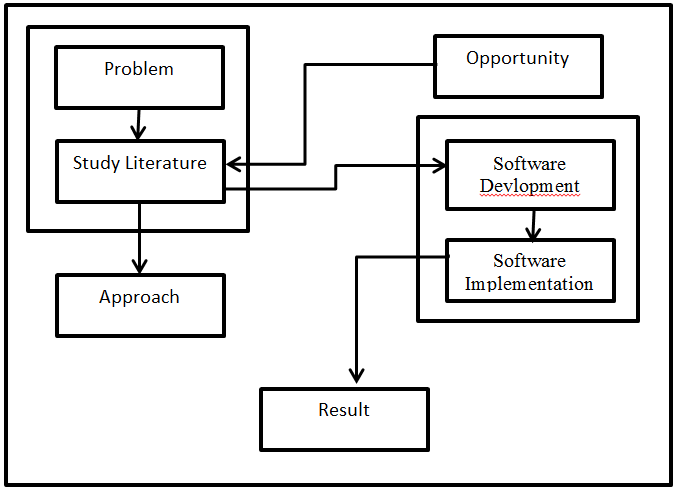
# BAB III

# ANALISIS DAN PERANCANGAN

Perancangan perangkat lunak yang akan dibuat dala skripsi ini akan menggunakan model pengembangan perangkat lunak *sequential linear.* Model pengembangan ini terdiri dari tahap analisis kebutuhan perangkat lunak, tahap desain, tahap pengkodean, tahap pengujian, dan tahap pemeliharaan. Tahap analisis kebutuhan perangkat lunak, tahap desain akan dibahas pada bab 3 ini, sedangkan tahap pengujian akan dibahas pada bab 4.

## Kerangka Kerja Konseptual

Gambar 3.1 di bawah ini merupakan gambaran mengenai kerangka kerja secara konseptual dalam penelitian kali ini yang disertai dengan penjelasannya.



**Gambar 3.1** Kerangka Kerja Konseptual

1. *Problem*

Masalah yang terjadi dalam penelitian ini adalah banyaknya data-data yang tersebar dalam dunia maya, khususnya yang berbentuk artikel berita dan belum diklasifikasikan sesuai topik yang dibahasnya atau masih menggunakan sistem klasifikasi yang manual. Data atau artikel yang belum terklasifikasikan membuat *user* mengalami kesulitan dalam mencari artikel dengan topik tertentu dan akan membutuhkan waktu yang cukup lama pada proses pencariannya.

1. *Study Literature*

*Study literature* dilakukan untuk membantu memecahkan permasalahan yang terjadi, yaitu dengan mencari data-data yang dibutuhkan untuk memecahkan permasalahan tersebut. Data-data yang dibutuhkan tersebut dicari melalui buku, *paper* dan jurnal penelitian, artikel, maupun melalui artikel *website* yang berhubungan dengan penelitian.

1. *Opportunity*

*Opportunity* atau peluang untuk melalukan penelitian ini karena belum adanya aplikasi yang dapat mengklasifikasikan data, khususnya artikel berita yang berasal dari situs berita Indonesia resmi secara otomatis.

1. *Approach*

Setelah melihat adanya permasalahan dan peluang untuk membuat aplikasi klasifikasi artikel berita internet berbasis *web*, maka didapatkan metode yang cocok untuk digunakan dalam penelitian.

1. *Software Development*

Model pengembangan perangkat lunak yang akan digunakan dalam membuat aplikasi ini yaitu model *sequential linear.* Tahap pengembangan perangkat lunak dengan model *sequential linear* terdiri dari lima tahap [7], yaitu:

1. Tahap analisis kebutuhan perangkat lunak *(system analysis)* adalah tahap untuk mengumpulkan kebutuhan perangkat lunak yang dilakukan dengan intensif.
2. Tahap desain *(system design)* adalah tahap dimana kebutuhan perangkat lunak yang sebelumnya didapatkan pada tahap analisis akan digambarkan ke dalam desain perangkat lunak.
3. Tahap pengkodean *(coding)* adalah tahap menerjemahkan desain perangkat lunak yang dibuat pada tahap desain ke dalam kode-kode bahasa pemrograman.
4. Tahap pengujian *(test)* adalah tahap untuk menguji fungsionalitas perangkat lunak untuk memastikan bahwa setiap bagian dari perangkat lunak tersebut berjalan sebagaiman mestinya.
5. Tahap pemeliharaan *(maintenance)* atau tahap pendukug *(support)* adalah tahap untuk pemeliharaan terhadap perangkat lunak yang sudah dikirimkan kepada *user* jika akan dilakukan perubahan atau perbaikan terhadap bagian perangkat lunak yang tidak berjalan sesuai keinginan *user*.
6. *Software Implementation*

*Software implementation* dilakukan dengan mengambil artikel dari beberapa *website* berita *online*, kemudian elemen *web* yang berupa *plaintext* (teks) akan dihitung nilainya untuk menentukan kategori yang tepat sesuai dengan topik yang dibahas pada artikel tersebut.

1. *Result*

Hasil akhir dari proses di atas adalah sebuah aplikasi yang dapat mengklasifikasikan artikel berita *online* secara otomatis dengan menggunakan metode *k-nearest neighbor* dan *naïve bayes* berbasis *website.*

## Analisis Sistem

Tahap analisis sistem menjelaskan mengenai analisis dan perancangan aplikasi yang akan dikembangkan. Proses analisis sistem terdiri dari analisis kebutuhan *non*-fungsional, kebutuhan fungsional, dan analisis mengenai metode yang akan digunakan dalam proses pengembangan aplikasi.

### Analisis Kebutuhan Non-Fungsional

Kebutuhan non-fungsional dalam pembuatan perangkat lunak ini terdiri dari tiga hal, yaitu kebutuhan perangkat keras *(hardware)*, kebutuhan perangkat lunak *(software),* dan analisis pengguna aplikasi.

1. Kebutuhan Perangkat Keras *(Hardware)*

Perangkat keras yang akan digunakan dalam pengembangan aplikasi menggunakan komputer dengan spesifikasi minimum sebagai berikut:

1. Intel Pentium Core 2 Duo 2.26 GHz;
2. RAM 2 GB;
3. *Harddisk* 40 GB;
4. *Monitor, keyboard, mouse* sebagai peralatan antarmuka.
5. Kebutuhan Perangkat Lunak *(Software)*

Kebutuhan perangkat lunak *(software)*

1. Sistem operasi *Windows* (Windows XP atau yang lebih tinggi);
2. *Software editor,* seperti *Sublime Text 2, notepad++;*
3. XAMPP *packet* versi 1.7.3.0yang terdiri dari applikasi *web server* versi*,* yaitu Apache versi 2.2.14 dan aplikasi *database server,* yaitu MySQL versi 5.1.41;
4. Bahasa pemrograman PHP versi 5.3.1;
5. Perangkat lunak pemodelan sistem, seperti *StarUML* versi 5.0.2.1570
6. Perangkat lunak pemodelan *user interface,* seperti *Balsamiq Mockups*
7. Analisis Kebutuhan Pengguna *(Client)*

Karakteristik pengguna yang akan berperan dalam penggunaan aplikasi ini yaitu hanya sebagai pemakai *(user)* saja. Pengguna hanya memerlukan perangkat komputer dengan perangkat lunak Google Chrome atau Mozilla Firefox yang tersambung dengan internet.

### Analisis Kebutuhan Fungsional

Analisis kebutuhan fungsional berisi mengenai deskripsi dan fitur-fitur yang terdapat pada aplikasi klasifikasi artikel berita *online* ini. Analisis kebutuhan fungsional meliputi deskripsi global dari aplikasi yang akan dikembangkan, arsitektural sistem yang akan dikembangkan, dan penggunaan metode yang akan diterapkan dalam aplikasi tersebut.

1. Deskripsi Global Aplikasi

Aplikasi klasifikasi artikel berita *online* merupakan aplikasi berbasis *website* yang berfungsi untuk mengklasifikasikan artikel berita yang berasal dari beberapa situs berita Indonesia. Aplikasi ini dikembangkan dengan tujuan untuk mempermudah dalam mengklasifikasikan data-data yang berada di dunia maya, khususnya artikel berita *online.*

1. Arsitektur Sistem

Arsitektur sistem dari aplikasi pengklasifikasian artikel berita *online* secara garis besar digambarkan pada Gambar 3.2 di bawah ini.

|  |
| --- |
|  |

**Gambar 3.2** Arsitektur Sistem

Arsitektur sistem yang dibangun melibatkan beberapa proses yang saling terhubung satu sama lain, sehingga menjadikan aplikasi yang dapat memberikan informasi mengenai kategori artikel berita *online* yang tepat. Berikut penjelasan masing-masing proses yang terjadi dalam arsitektur sistem di atas.

*Visitor*

*User* atau pengguna merupakan orang yang akan menggunakan *web* yang akan dibuat untuk mengetahui informasi yang berasal dari artikel di internet.

*Website*

*Website* disini merupakan sebuah halaman yang akan diambil datanya untuk diolah dan ditemukan kategorinya.

*Preprocessing*

*Preprocessing* merupakan tahap untuk mempersiapkan teks menjadi data yang akan diolah ke tahapan selanjutnya [11]. Ada empat tahapan dalam proses *preprocessing,* yaitu tahap *case folding,* tahap *tokenizing,* tahap *filtering*, dan tahap *stemming.*

Tahap *case folding* merupakan tahapan dimana teks dalam dokumen akan diubah menjadi huruf kecil semua serta menghilangkan semua karakter dalam dokumen selain huruf a-z, kecuali karakter pemisah kata seperti spasi, tab, maupun *newline* (baris baru). Tahap *tokenizing* merupakan tahap pemotongan *string* setiap kata yang menyusun dokumen tersebut. Pemisahan kata ini berdasarkan pemisah kata diantara dua kata tersebut, seperti spasi, tab, maupun *newline* (baris baru). Tahap *filtering* merupakan tahap untuk menghilangkan *stopword.* *Stopword* merupakan kata yang sering muncul dalam dokumen tersebut tetapi artinya tidak deskriptif dan tidak memiliki arti yang terkait dengan tema tertentu [11]. Tahap *stemming* merupakan tahap pencarian kata dasar dari suatu kata atau pengembalian kata yang berimbuhan ke dalam kata dasarnya.

*K-Nearest Neighbour Clustering*

Algoritma *K-Nearest Neighbour* digunakan untuk melakukan pengklasteran pada teks dalam dokumen yang sebelumnya sudah dilakukan proses *preprocessing.* Proses pengklasteran menggunakan algoritma KNN ini dilakukan berdasarkan data yang jaraknya terdekat dengan teks tersebut.

Algoritma *Naive Bayes*

Algoritma *Naive Bayes* menggunakan perhitungan probabilitas, yaitu dengan memprediksi peluang di masa depan berdasarkan pengalaman sebelumnya [5]. Algoritma ini akan mengklasifikasikan suatu objek dengan mengasumsikan ada atau tidaknya ciri tertentu dari suatu kelas yang satu dan tidak ada hubungannya dengan ciri-ciri pada kelas yang lain.

1. Arsitektur Pengolahan Data

|  |
| --- |
|  |

**Gambar 3.3** Arsitektur Sistem Pengolah Data

Gambar 3.3 di ata menggambarkan arsitektur sistem pengolah data yang dimulai dari mengambil konten halaman *website* berita *online* berupa artikel berita kemudian disimpan ke dalam *database* dalam sebuah *server* yang nantinya dapat diakses oleh pengguna melalui halaman *website* aplikasi kategori berita *online.*

### Analisis Masalah Penentuan Kategori Berita

Untuk menentukan kategori berita dilakukan proses perhitungan matematis yang sudah sudah dijelaskan pada bab landasan teori di atas, yaitu menggunakan metode *k-nearest neighbour* dan *naïve bayes.* Gambar 3.4 di bawah ini menjelaskan mengenai fase penentuan kategori berita, yaitu:

|  |
| --- |
|  |

**Gambar 3.4** Fase Penentuan Kategori

1. ***Scraping Website***

*Scraping website* merupakan tahap untuk mengambil isi dari halaman *web* secara spesifik. Pada tahap *scraping* akan diambil isi *web* berupa *link website, link image,* dan isi artikel berita yang pada tahap selanjutnya akan dilakukan proses perhitungan untuk mendapatkan kategori yang sesuai dengan topik berita. Contoh dari isi artikel berita ditunjukan pada Gambar 3.5 di bawah ini:

|  |
| --- |
| Ketua Komisi III DPR Bambang Soesatyo, membenarkan adanya operasi tangkap tangan oleh KPK terhadap salah satu anggota Komisi III.  Anggota tersebut, diduga dari Fraksi Partai Demokrat. Namun Bambang memastikan tidak terkait kegiatan di komisi.  "Saya sudah cek, tidak terkait dengan tupoksi Komisi III. Lebih kepada kegiatan Banggar," kata Bambang, saat dikonfirmasi, Rabu 29 Juni 2016.  Bambang mengaku sedih dengan penangkapan rekan sejawat di komisinya tersebut. Bambang yang juga Ketua DPP Golkar, menyebut Putu sebagai sosok yang humoris.  "Kami semua di Komisi III sedih dan prihatin. Putu (I Putu Sudiartana) adalah sahabat yang baik dan humoris. Nggak ada Putu nggak rame," tutur Bambang.  Saat ini, KPK juga sudah menyegel ruang anggota tersebut di DPR.  Tim Satuan Tugas KPK dikabarkan telah menangkap seorang anggota DPR karena diduga terkait dengan tindak pidana korupsi. Anggota dewan itu disebut-sebut berasal dari Komisi lll DPR Fraksi Partai Demokrat berinisial IPS.  Saat dikonfirmasi, Wakil Ketua KPK, Laode Muhammad Syarif tidak menampiknya. Namun menurut Syarif, penjelasan lebih detail akan diberikan pada konferensi pers nanti.  "Tunggu sore saja detailnya," kata Syarif. |

**Gambar 3.5** Contoh Artikel Berita

1. ***Text Preprocessing***

*Text preprocessing* merupakan tahap untuk mempersiapkan dokumen teks menjadi data yang akan diolah pada tahap selanjutnya. *Text preprocessing* terdiri dari beberapa tahapan, diantaranya pemecahan kalimat, *case folding, tokenizing, filtering,* dan *stemming.*

**Pemecahan Kalimat**

Pemecahan kalimat merupakan tahap untuk memecah *string* dokumen teks menjadi beberapa kalimat dengan memperhatikan tanda pemisah kalimat seperti titik, tanda tanya, dan tanda seru sebagai *delimeter* pemisah kalimat. Contoh pemecahan kalimat ditunjukan pada Tabel 3.1 di bawah ini:

**Tabel 3.1** Pemecahan Kalimat

|  |  |
| --- | --- |
| **No** | **Kalimat** |
| 1. | Ketua Komisi III DPR Bambang Soesatyo, membenarkan adanya operasi tangkap tangan oleh KPK terhadap salah satu anggota Komisi III. |
| 2. | Anggota tersebut, diduga dari Fraksi Partai Demokrat. Namun Bambang memastikan tidak terkait kegiatan di komisi. |
| 3. | "Saya sudah cek, tidak terkait dengan tupoksi Komisi III. Lebih kepada kegiatan Banggar," kata Bambang, saat dikonfirmasi, Rabu 29 Juni 2016. |
| 4. | Bambang mengaku sedih dengan penangkapan rekan sejawat di komisinya tersebut. Bambang yang juga Ketua DPP Golkar, menyebut Putu sebagai sosok yang humoris. |
| 5. | "Kami semua di Komisi III sedih dan prihatin. Putu (I Putu Sudiartana) adalah sahabat yang baik dan humoris. Nggak ada Putu nggak rame," tutur Bambang. |
| 6. | Saat ini, KPK juga sudah menyegel ruang anggota tersebut di DPR. |
| 7. | Tim Satuan Tugas KPK dikabarkan telah menangkap seorang anggota DPR karena diduga terkait dengan tindak pidana korupsi. Anggota dewan itu disebut-sebut berasal dari Komisi III DPR Fraksi Partai Demokrat berinisial IPS. |
| 8. | Saat dikonfirmasi, Wakil Ketua KPK, Laode Muhammad Syarif tidak menampiknya. Namun menurut Syarif, penjelasan lebih detail akan diberikan pada konferensi pers nanti. |
| 9. | "Tunggu sore saja detailnya," kata Syarif. |

***Case Folding***

*Case folding* merupakan tahap untuk mengubah semua huruf menjadi huruf kecil *(lower case)* dan menghilangkan semua simbol yang ada dalam teks kecuali huruf a-z. Tabel 3.2 di bawah ini menunjukan hasil *case folding* dari Tabel 3.1 di atas.

**Tabel 3.2** *Case Folding*

|  |  |
| --- | --- |
| **No** | **Kalimat** |
| 1. | ketua komisi iii dpr bambang soesatyo membenarkan adanya operasi tangkap tangan oleh kpk terhadap salah satu anggota komisi iii |
| 2. | anggota tersebut diduga dari fraksi partai demokrat namun bambang memastikan tidak terkait kegiatan di komisi |
| 3. | saya sudah cek tidak terkait dengan tupoksi komisi iii lebih kepada kegiatan banggar kata bambang saat dikonfirmasi rabu juni |
| 4. | bambang mengaku sedih dengan penangkapan rekan sejawat di komisinya tersebut bambang yang juga ketua dpp golkar menyebut putu sebagai sosok yang humoris |
| 5. | kami semua di komisi iii sedih dan prihatin putu i putu sudiartana adalah sahabat yang baik dan humoris nggak ada putu nggak rame tutur bambang |
| 6. | saat ini kpk juga sudah menyegel ruang anggota tersebut di dpr |
| 7. | tim satuan tugas kpk dikabarkan telah menangkap seorang anggota dpr karena diduga terkait dengan tindak pidana korupsi anggota dewan itu disebut sebut berasal dari komisi iii dpr fraksi partai demokrat berinisial ips |
| 8. | saat dikonfirmasi wakil ketua kpk laode muhammad syarif tidak menampiknya namun menurut syarif penjelasan lebih detail akan diberikan pada konferensi pers nanti |
| 9. | tunggu sore saja detailnya kata syarif |

***Tokenizing***

*Tokenizing* merupakan tahap untuk memecah kata-kata dalam teks yang menyusun kalimat/*string.* Pemotongan kata dilakukan dengan melihat pemisah kata berupa spasi sebagai *delimeter* pemisah kata. Tabel 3. 3 di bawah ini menunjukan hasil dari *tokenizing.*

**Tabel 3.3** *Tokenizing*

|  |  |  |  |
| --- | --- | --- | --- |
| **Kata** | **Kata** | **Kata** | **Kata** |
| ketua | komisi | iii | dpr |
| bambang | soesatyo | membenarkan | adanya |
| anggota | tangkap | tangan | oleh |
| kpk | terhadap | salah | satu |
| tersebut | diduga | dari | fraksi |
| partai | demokrat | namun | memastikan |
| tidak | terkait | kegiatan | di |
| saya | sudah | cek | dengan |
| tupoksi | lebih | kepada | banggar |
| kata | saat | dikonfirmasi | rabu |
| juni | mengaku | sedih | penangkapan |
| rekan | sejawat | komisinya | yang |
| juga | dpp | golkar | menyebut |
| putu | sebagai | sosok | humoris |
| kami | semua | dan | prihatin |
| i | sudiartana | adalah | sahabat |
| baik | nggak | ada | rame |
| tutur | ini | menyegel | ruang |
| tim | satuan | tugas | dikabarkan |
| telah | menangkap | seorang | karena |
| tindak | pidana | korupsi | dewan |
| itu | disebut | sebut | berasal |
| berinisial | ips | wakil | laode |
| muhammad | syarif | menampiknya | menurut |
| penjelasan | detail | akan | diberikan |
| pada | konferensi | pers | nanti |
| tunggu | sore | saja | detailnya |

***Filtering***

*Filtering* merupakan tahap penyaringan kata dengan menghilangkan kata-kata berupa *stopword* sepeti “yang”, “itu”, “ini”, “adalah”, dan yang lainnya. Penghilangan kata *stopword* dilakukan dengan melakukan pengecekan pada kamus *stopword.* Jika terdapat kata yang sama dengan daftar di kamus *stopword* maka kata tersebut akan dihilangkan. Tabel 3.4 dibawah ini menunjukan kata-kata yang dianggap ke dalam daftar *stopword,* yaitu:

**Tabel 3.4** Kata *Stopword*

|  |  |  |  |
| --- | --- | --- | --- |
| **Kata** | **Kata** | **Kata** | **Kata** |
| adanya | oleh | terhadap | salah |
| satu | tersebut | diduga | dari |
| namun | tidak | terkait | di |
| saya | sudah | dengan | lebih |
| kepada | kata | saat | rabu |
| juni | mengaku | yang | juga |
| menyebut | sebagai | kami | semua |
| dan | adalah | baik | nggak |
| ada | tutur | ini | telah |
| seorang | karena | itu | menurut |
| akan | pada | nanti | saja |

Tabel 3.5 di bawah ini merupakan hasil dari *filtering.*

**Tabel 3.5** *Filtering*

|  |  |  |  |
| --- | --- | --- | --- |
| **Kata** | **Kata** | **Kata** | **Kata** |
| ketua | komisi | iii | dpr |
| bambang | soesatyo | membenarkan | anggota |
| tangkap | tangan | kpk | fraksi |
| partai | demokrat | memastikan | kegiatan |
| cek | tupoksi | banggar | dikonfirmasi |
| sedih | penangkapan | rekan | sejawat |
| komisinya | dpp | golkar | putu |
| sosok | humoris | prihatin | i |
| sudiartana | sahabat | rame | menyegel |
| ruang | tim | satuan | tugas |
| dikabarkan | menangkap | tindak | pidana |
| korupsi | dewan | disebut | sebut |
| berasal | berinisial | ips | wakil |
| laode | muhammad | syarif | menampiknya |
| penjelasan | detail | diberikan | konferensi |
| pers | tunggu | sore | detailnya |

***Stemming***

*Stemming* merupakan tahap untuk mencari *root* kata dari setiap kata yang menyusun kalimat, yaitu dengan mengembalikan kata berimbuhan menjadi kata dasar. Proses *stemming* dilakukan dengan mencocokkan kata di kamus kata dasar dan membuang imbuhan yang menyusun kata tersebut. Tabel 3.6 di bawah ini menunjukan hasil dari proses *stemming.*

**Tabel 3.6** *Stemming*

|  |  |  |  |
| --- | --- | --- | --- |
| **Kata** | **Kata** | **Kata** | **Kata** |
| ketua | komisi | iii | dpr |
| bambang | soesatyo | benar | anggota |
| tangkap | tangan | kpk | fraksi |
| partai | demokrat | pasti | giat |
| cek | tupoksi | banggar | konfirmasi |
| sedih | tangkap | rekan | jawat |
| komisi | dpp | golkar | putu |
| sosok | humoris | prihatin | i |
| sudiartana | sahabat | rame | segel |
| ruang | tim | satu | tugas |
| kabar | tangkap | tindak | pidana |
| korupsi | dewan | sebut | asal |
| inisial | ips | wakil | laode |
| muhammad | syarif | tampik | jelas |
| detail | beri | konferensi | pers |
| tunggu | sore |  |  |

1. **Pembobotan TF-IDF**

Pembobotan tf-idf dilakukan setelah tahap *preprocessing* selesai diproses. Pembobotan dilakukan berdasarkan penomoran dokumen, dimana tiap kata yang menyusun dokumen akan dihitung bobotnya berdasarkan banyaknya kata dari tiap dokumen. Tabel 3.8 menampilkan pembobotan tf-idf dengan beserta 15 data latih yang sudah diklasifikasikan ke dalam kategori Politik (D2 , D3 dan D4), Olahraga (D5, D6, dan D7), Pendidikan (D8, D9, dan D10), Otomotif (D11, D12,

dan D13), dan Umum (D14, D15, dan D16). *Term* yang terdapat pada Tabel 3.6 merupakan kata-kata yang selalu ada dan sering muncul dari setiap dokumen. Berikut ini merupakan dokumen data latih yang ditunjukkan dalam Tabel 3.7.

**Tabel 3.7** Daftar Data Latih

|  |  |
| --- | --- |
| **Dok** | **Isi Dokumen** |
| D2 | Rancangan revisi atas Peraturan Pemerintah (PP) Nomor 99 Tahun 2012 – tentang perubahan kedua atas PP Nomor 32 tahun 1999 tentang Syarat dan Tata Cara Pelaksanaan Hak Warga Binaan Pemasyarakatan – menyebutkan bahwa ketentuan justice collabolator (JC) sebagai syarat remisi bagi pelaku tindak pidana korupsi, terorisme, dan narkotika, akan dihilangkan.  Implikasinya, terpidana kasus tersebut bisa mendapat remisi, hanya dengan dua syarat pokok, yakni berkelakuan baik dan telah menjalani sepertiga masa pidananya.  Terkait hal itu, Wakil Presiden (Wapres) Jusuf Kalla (JK) setuju atas rencana perubahan pemerintah yang sedang digodok Kementerian Hukum dan Hak Asasi Manusia. Alasannya adalah faktor kemanusiaan.  "Semua orang yang telah dihukum, dipenjara tentu mempunyai juga sisi-sisi kemanusiaan. Apapun (kasusnya). Kami tidak membedakan lagi (kasus narapidana)," ujar JK di Kantor Wakil Presiden, Jalan Medan Merdeka Utara, Jakarta Pusat, Jumat 12 Agustus 2016.  Meski demikian, JK mengakui bahwa remisi tersebut akan sangat sulit untuk diberikan kepada koruptor. Namun menurutnya remisi bisa menjadi hal yang membuat koruptor dan pelaku tindak kejahatan lainnya bertobat.  "Walaupun tentu kami agak berat. Gunanya remisi sebenarnya, intinya agar mereka itu memperlihatkan disiplin, merasa katakanlah bertobat dari sisi moral atau berkelakuan baik. Itulah syarat remisi itu," kata JK.  Dia membandingkan jika narapidana dengan kasus pembunuhan yang menghilangkan nyawa orang saja masih diberikan remisi, lantas koruptor juga seharusnya memiliki hak tersebut. Perlakuan yang sama harus diberikan agar negara tidak dianggap diskriminatif.  "Kalau pembunuh saja bisa (mendapat remisi), kemudian koruptor tidak bisa diberikan reward karena disiplinnya karena kelakuan baiknya, tentu juga kita terjadi diskriminatif," lanjut JK.  Apalagi kata JK, selama ini banyak koruptor yang telah dijatuhi hukuman berat dengan masa kurungan yang lama.  "Walaupun kita tahu korupsi itu kejahatan besar tapi narkoba dan pembunuhan juga kejahatan besar juga efeknya terhadap negara. Jadi kita lihat dari sisi kemanusiaan. Kalau dia tobat berkelakuan baik, makin baik dia punya perilaku. Bukan lihat lagi dari sisi apa yang dia buat karena ringan beratnya hukuman kan sudah ada UU-nya, sudah ada pengadilannya," kata Wapres.  Oleh karena itu perilaku saat dan setelah narapidana menjalani hukuman kata JK adalah pertimbangan penting. |
| D3 | Partai Demokrasi Indonesia Perjuangan (PDIP) dikabarkan bakal membahas bakal calon gubernur (cagub) DKI Jakarta hari ini. Setelah sebelumnya, Partai Moncong Putih juga telah melakukan rapat internal di Kantor Dewan Pimpinan Pusat (DPP), PDI Perjuangan dua hari lalu.  Namun, Sekretaris Jenderal PDIP, Hasto Kristiyanto yang siang ini baru saja meninggalkan kediaman Ketua Umum PDIP, Megawati Soekarnoputri di Jalan Teukur Umar, Menteng, Jakarta Pusat, membantahnya. |

**Tabel 3.7** Daftar Data Latih (Lanjutan)

|  |  |
| --- | --- |
| **Dok** | **Isi Dokumen** |
|  | "Hari ini rapat rutin partai, Kamis dan Jumat persoalan internal seperti pilkada, kaderisasi, penegakan disiplin partai. Jumat, lebih pada kondisi eksternal. Jadi, hari ini tidak ada pembahasan Pilkada DKI," jelas Hasto di Jakarta, Jumat 12 Agustus 2016.  Lebih lanjut, Hasto menjelaskan, kondisi eksternal yang dia maksud adalah hal strategis soal pengelolaan pemerintahan.  "Eksternal itu hal-hal strategis, misal menerima undangan Lemhannas (Lembaga Ketahanan Nasional), undangan-undangan (ke partai) yang masuk kami sampaikan. Skenario sudah ada tinggal mengambil keputusan akhir," kata dia lagi.  Hasto juga membantah informasi yang beredar tentang kedatangan Wali Kota Surabaya Tri Rismaharini Jakarta hari ini.  "Dan, terkait Risma yang akan datang ke Ibu Kota itu tidak benar," katanya.  Sebelumnya, Hasto juga membantah bahwa PDIP akan mengumumkan nama bakal calon gubernur yang bakal diusung di Pilkada 2017 dalam waktu dekat."Tidaklah, kami mencari momentum yang baik terhadap pengumuman calon kepala daerah tersebut," kata Hasto. |
| D4 | Ketua DPD Hanura DKI Jakarta Muhammad "Ongen" Sangaji mengatakan akan ada dua partai politik lagi yang bergabung dengan koalisi pendukung Gubernur Basuki Tjahaja Purnama atau Ahok maju lagi dalam Pilkada DKI 2017.  Saat ini sudah ada tiga partai politik yang mendukung Ahok, yakni Hanura, Golkar, dan Nasdem.  "Saya sudah sampaikan kemarin kan, dalam waktu menjelang 10 hari ini ada dua partai yang bergabung dan itu pasti," kata Ongen di Gedung DPRD DKI Jakarta, Jumat (12/8/2016).  Namun, Ongen tak menyebutkan partai mana yang akan bergabung dalam koalisi pendukung Ahok. Menurut dia, saat ini kedua partai itu menjadi informan bagi koalisi pendukung Ahok untuk memantau pergerakan "Koalisi Kekeluargaan".  "Pokoknya biarkan kami titip mereka di dalam 10 hari biar lebih tahu apa yang terjadi di antara tujuh partai itu," kata Ongen.  Tujuh partai dalam Koalisi Kekeluargaan adalah PDI-P, Gerindra, PKS, PPP, Demokrat, PKB, dan PAN.  Koalisi itu baru diputuskan di tingkat DPD dan belum ada perjanjian hitam di atas putih untuk membuat Koalisi Kekeluargaan. |
| D5 | Ganda putra Indonesia Hendra Setiawan/Mohammad Ahsan tampil buruk sehingga kalah dari Endo Hiroyuki/Kenichi Hayakawa. Hendra/Ahsan memang underperform.  Hendra/Ahsan secara mengejutkan kalah dari pasangan Jepang itu dalam duel tiga gim 17-21, 21-16, 14-21. Padahal, di sembilan pertemuan sebelumnya, mereka selalu berhasil memetik kemenangan.  Satu hal yang menjadi sorotan adalah seringnya service foul yang dilakukan Hendra. Tercatat sembilan kali pelanggaran itu terjadi yang artinya poin cuma-cuma untuk lawan.  "Hendra/Ahsan memang tampil underperformed. Ada rasa nervous, walaupun sebagai pemain senior seharusnya nggak boleh nervous, tapi ini kan olimpiade, kejuaraan besar. Hendra/Ahsan kelihatan sering terburu-buru dan kurang tenang, mainnya tidak lepas dan banyak melakukan kesalahan sendiri," ujar Kepala Pelatih Ganda Putra PP PBSI Herry Iman Pierngadi. |

**Tabel 3.7** Daftar Data Latih (Lanjutan)

|  |  |
| --- | --- |
| **Dok** | **Isi Dokumen** |
|  | "Hari ini Hendra banyak melakukan kesalahan di servis. Sebetulnya kami sudah mengantisipasi hal ini, waktu di karantina. Bahkan ada latihan servis khusus untuk Hendra. Namun kalau lagi tanding, yang tahu perasaannya kan Hendra sendiri."  "Ahsan sebetulnya terpengaruh, mungkin dia mencoba untuk menutupi dengan dua tiga kali pukulan, tetapi lawan tidak mudah ditembus, akhirnya Ahsan membuat kesalahan sendiri," imbuh dia dalam rilis yang diterima dari PP PBSI  Situasi sulit dihadapkan pada Hendra/Ahsan. Keduanya wajib menang atas wakil China Chai Biao/Hong Wei. Akan tetapi, kemenangan pun akan menempatkan Hendra/Ahsan sebagai runner-up Grup D yang mana berpotensi melakoni laga berat karena berjumpa dengan juara dari grup-grup lain.  "Tentunya besok harus lebih maksimal dan ini tidak gampang, tekanan pasti tinggi. Siapa yang bisa mengatasi tekanan, dialah yang akan menang. Kami akan mengevaluasi penampilan Hendra/Ahsan dan fokus melawan Chai/Hong. Kalau menang, perjuangan di perempat final sebagai runner up memang lebih berat," demikian Herry. |
| D6 | Ganda campuran Tontowi Ahmad/Liliyana Natsir tak mengalami kesulitan berarti untuk meraup kemenangan kedua. Hasil ini meloloskan Tontowi/Liliyana ke perempatfinal.  Bermain di lapangan 1 Riocentro, Jumat (12/8/2016), Tontowi/Liliyana menghadapi pasangan Thailand Bodin Isara/Savitree Samitrapai di Grup C. Hanya dalam 32 menit, Tontowi/Liliyana menang dua gim langsung 21-11, 21-13.  Tontowi/Liliyana tampil begitu dominan di gim pertama. Keduanya melesat dengan keunggulan 16-4 sebelum memenangi laga 21-11.  Akan tetapi, gim kedua tidak berjalan mudah bagi Tontowi/Liliyana. Setelah sempat tertinggal, pasangan Thailand mampu meraup lima poin beruntun untuk berbalik unggul 11-6.  Meski demikian, Tontowi/Liliyana mampu menyusul bahkan berbalik unggul 14-13 setelah pengembalian lawan menyangkut di net.  Tontowi/Liliyana terus melesat dan sergapan Liliyana memastikan match point. Pasangan Indonesia akhirnya memenangi pertandingan usai smash keras Tontowi tidak dapat dikembalikan Samitrapai.  Di pertandingan terakhir Tontowi/Liliyana akan menghadapi pasangan Malaysia Chan Peng Soon/Goh Liu Ying untuk memperebutkan gelar juara grup C. |
| D7 | Pelaksana Tugas (Plt) Ketua Umum PSSI, Hinca Panjaitan, membeberkan agenda tim nasional Indonesia, baik tingkat U-19 maupun senior, untuk 2016. Hal itu disampaikan oleh Hinca di kantor PSSI di kawasan Kuningan, Jakarta, pada Kamis (11/8/2016).  Menurut Hinca, untuk tahun ini, timnas Indonesia hanya akan mengikuti dua turnamen internasional, yakni Piala AFF U-19 (untuk Timnas U-19) dan Piala AFF (untuk timnas senior).  Piala AFF U-19 akan dilangsungkan di Vietnam pada 11-24 September 2016. Adapun Piala AFF, fase penyisihan grup diselenggarakan di Myanmar dan Filipina pada 19 November hingga 17 Desember 2016.  Menurut Hinca, sebelum mengikuti kedua ajang tersebut, tim nasional dijadwalkan menjalani partai persahabatan yang semuanya digelar di Stadion Manahan, Solo.  Timnas U-19 akan melakukan laga uji coba melawan Filipina pada 20 Agustus 2016. Sementara itu, timnas senior bakal menghadapi Malaysia pada 6 September 2016.  "Kita akan melawan musuh bebuyutan, yakni Malaysia. Tak ramai rasanya kalau di Asia tak ada Indonesia dan Malaysia. Mudah-mudah tidak berubah lagi |

**Tabel 3.7** Daftar Data Latih (Lanjutan)

|  |  |
| --- | --- |
| **Dok** | **Isi Dokumen** |
|  | tempatnya, di Stadion Manahan Solo," ujar Hinca.  Sebelumnya, 24 dari 47 pemain sudah diseleksi untuk masuk timnas senior pada 9 dan 10 Agustus 2016 di Stadion Pakansari, Cibinong, Kabupaten Bogor.  Sisanya, 23 pemain masuk seleksi tahap kedua pada 16 dan 17 Agustus 2016. |
| D8 | Proyek pembangunan infrastruktur yang digalakkan pemerintah saat ini mulai diiringi dengan penyediaan sarjana teknik maupun profesi keinsinyuran.  Di Kota Semarang, Jawa Tengah, Jumat (12/8/2016), Universitas Katolik Soegijapranata mulai membuka program studi profesi insinyur. Program ini diperuntukkan khusus bagi lulusan sarjana teknik dengan masa studi selama satu tahun.  Program profesi insinyur merupakan mandataris dari Kementerian Riset dan Pendidikan Tinggi agar para insinyur Indonesia mempunyai keahlian dan sertifikat profesi sesuai bidangnya. Mereka selain dididik belajar teknik, mereka juga nantinya akan mendapat gelar profesi insinyur.  “Program prefesi insinyur ini paketnya satu tahun selesai. Syaratnya harus sarjana teknik dulu, karena saat ini kerja global dibutuhkan sertifikat keahlian khusus,” ujar Wakil Rektor Unika Augustina Sulastri, seusai kegiatan peresmian.  Pendidikan profesi insinyur hari ini resmi diluncurkan oleh Rektor Budi Widianarko. Masa kuliah di kampus tersebut akan dimulai pada 12 September 2016. Selain Unika, ada tiga kampus lain di Jateng yang mendapat tugas serupa. Yaitu Universitas Diponegoro, Universitas Sebelas Maret, dan Universitas Muhammadiah.  Sulastri mengatakan, para sarajana teknis dan calon insinyur ini akan digembleng materi pendidikan. Pendidikan di perguruan tinggi diperlukan sebagai prasyarat mengikuti sertifikasi profesi yang dikeluarkan oleh Persatuan Insinyur Indonesia (PII). Para calon insinyur yang hendak melakukan sertifikat harusnya mengikuti perkuliahan profesi terlebih dulu.  “Jadi sebelum ikut sertifikasi di PII harus studi sertifikat profesi dulu,” tambah dia.  “Prodi sertifikasi Profesi hanya bisa diselenggarakan perguruan tinggi untuk menyiapkan tenaga profesional di pasar nasional, hingga ASEAN. Jadi ada kemudahan sertifikasi, tapi ada bekal pendidikan,” ujarnya.  Namun demikian, sebelum mengambil pendidikan profesi ini, calon mahasiswa harus mempunyai bekal minimal dua tahun bekerja di bidang keinsinyuran. Kampus tidak menerima para sarjana yang baru saja lulus atau fresh graduate.  Rektor Unika Budi Widinarko mengatakan, pembukaan prodi sebagai mandat dari pemerintah pada institusi perguruan tinggi. Setidaknya ada 25 perguruan tinggi negeri dan 15 perguruan tinggi swasta yang ditugaskan pemerintah untuk meningkatkan kompetensi profesi keinsinyuran. |
| D9 | Pelajar SMA IPEKA, Wilson Gomarga (18), berhasil membawa pulang medali emas pada kompetisi International Biology Olympiad (IBO) ke-27 di Hanoi, Vietnam, pada 16-23 Juli 2016 lalu. Wilson menyingkirkan 253 peserta dari 68 negara.  Pada ajang tahunan olimpiade internasional tersebut peserta asal Indonesia berasal dari empat sekolah. Indonesia berhasil mendapatkan 4 medali pada ajang IBO ke-27 ini.  "Ini pertama kalinya saya mengikuti IBO, dan tentunya senang dapat menyumbangkan medali emas mewakili Indonesia," ujar Wilson, Senin (1/8/2016).  Kiprah Wilson di ajang international berawal dari keikutsertaannya pada Olimpiade Sains Nasional (OSN) ke-14 tahun lalu di Yogyakarta. Wilson berhasil membawa pulang medali emas untuk bidang studi yang sama dan mendapat nilai praktik dan teori tertinggi.  Setelah itu, Wilson dipanggil lagi untuk mengikuti seleksi pada Pelatnas IBO 2016. Saat itu hanya tersisa 9 orang untuk mengikuti seleksi masuk IBO 2016. |

**Tabel 3.7** Daftar Data Latih (Lanjutan)

|  |  |
| --- | --- |
| **Dok** | **Isi Dokumen** |
|  | Wilson pun terpilih masuk pelatnas empat besar mewakili Indonesia di IBO 2016.  Sejak SMP, pelajar kelahiran Jakarta, 22 Juni 1998, itu mengaku sudah menyukai pelajaran biologi. Menurut dia, biologi bukanlah pelajaran hafalan, tapi lebih mirip novel yang harus dimengerti konsepnya supaya enak dan mudah dipahami.  "Saya ingin mendapatkan kesempatan pendidikan lebih baik lagi ke depan. Saya akan melanjutkan studi yang berkaitan dengan biologi dan kedokteran serta mendalami penerapan yang terdapat dalam bidang ilmu tersebut," kata penyuka olah raga bola tenis meja ini.  Sejauh ini, lanjut dia, pihak sekolah mendukungnya mengikuti ajang tersebut. Para guru membantunya ketika membutuhkan alat-alat peraga untuk berlatih di rumah.  "Saya dipinjami alat dan sebagainya untuk berlatih," tambahnya.  Kepala Bidang Pengembangan Pendidikan IPEKA, Theo Kurniadi, mengaku memang mendukung murid-muridnya dan mengajak para guru untuk memberikan pendampingan dan bimbingan kepada siswa-siswi.  "Pada prinsipnya harus lakukan yang terbaik," kata Theo.  Adapun International Biology Olympiad (IBO) merupakan ajang kompetisi bagi para siswa setingkat SMA dari seluruh dunia yang meliputi tes praktikum dan tes teori di bidang biologi. Kegiatan ini adalah ajang kompetisi tahunan yang dilaksanakan tiap Juli. Setiap negara peserta diwakili empat siswa yang diseleksi secara nasional. Indonesia mengikuti kompetisi ini sejak 2000 di Antalya, Turki. |
| D10 | Menteri Pendidikan dan Kebudayaan Muhadjir Effendy menggagas sistem "full day school" untuk pendidikan dasar (SD dan SMP), baik negeri maupun swasta. Alasannya agar anak tidak sendiri ketika orangtua mereka masih bekerja.  "Dengan sistem full day school ini secara perlahan anak didik akan terbangun karakternya dan tidak menjadi liar di luar sekolah ketika orangtua mereka masih belum pulang dari kerja," kata Mendikbud di Universitas Muhammadiyah Malang (UMM), Minggu (7/8/2016).  Menurut dia, kalau anak-anak tetap berada di sekolah, mereka bisa menyelesaikan tugas-tugas sekolah sampai dijemput orangtuanya seusai jam kerja.  Selain itu, anak-anak bisa pulang bersama-sama orangtua mereka sehingga ketika berada di rumah mereka tetap dalam pengawasan, khususnya oleh orangtua.  Untuk aktivitas lain misalnya mengaji bagi yang beragama Islam, menurut Mendikbud, pihak sekolah bisa memanggil guru mengaji atau ustaz dengan latar belakang dan rekam jejak yang sudah diketahui. Jika mengaji di luar, mereka dikhawatirkan akan diajari hal-hal yang menyimpang.  Menyinggung penerapan full day school dalam pendidikan dasar tersebut, mantan Rektor UMM itu mengatakan bahwa hal itu saat ini masih terus disosialisasikan di sekolah-sekolah, mulai di pusat hingga di daerah.  "Nantinya memang harus ada payung hukumnya, yakni peraturan menteri (permen). Namun, untuk saat ini masih sosialisasi terlebih dahulu secara intensif," ujarnya.  Sementara itu, ketika berbicara di hadapan ratusan kader Muhammadiyah Kota Malang, Muhadjir mengatakan, dirinya akan berupaya merestorasi pendidikan dasar dan menengah (SD-SMP), termasuk pendidikan karakter bagi anak didik. Selain itu, ia juga akan membenahi kebijakan-kebijakan yang berkaitan dengan profesionalisme para pendidik.  "Saya tidak akan mengutak-atik masalah sertifikasi guru. Namun, harapan saya, profesionalisme seorang guru juga harus ditingkatkan terus. Jangan ada |

**Tabel 3.7** Daftar Data Latih (Lanjutan)

|  |  |
| --- | --- |
| **Dok** | **Isi Dokumen** |
|  | guru yang tidak layak, tetapi tetap saja menuntut sertifikasi, bahkan prosesnya minta dipermudah," kata Mendikbud.  Menyinggung pendidikan di jenjang SMA dan SMK, Muhadjir mengatakan akan mencari formulasi yang tepat karena tidak semua lulusan SMA melanjutkan tahap pendidikan ke perguruan tinggi, alih-alih memilih untuk bekerja. Namun, karena tidak memiliki keterampilan dan keahlian, mereka akhirnya tidak bisa apa-apa di dunia kerja.  Walau demikian, lulusan SMK pun tidak semuanya langsung bekerja. Ada yang tetap melanjutkan tahap pendidikan ke perguruan tinggi. Meski mereka memiliki keterampilan sesuai minat yang diambil di SMK, jika kualitasnya tidak ditingkatkan dan memiliki keahlian yang memadai, mereka akan tergusur oleh tenaga kerja asing yang memiliki sertifikasi internasional.  "Kondisi ini yang akan kami carikan solusi agar kesenjangan dalam pendidikan bisa diminimalkan," ujarnya. |
| D11 | Aura klasik pada W800 sangat kuat. Itu bisa dilihat dari garis bodi yang membulat, seperti tangki, lampu, spakbor hingga speedometer analog. Rem depannya sudah menggunakan cakram, sementara belakang masih tromol. Kawasaki W800 mengandalkan mesin 4-Tak Vertical Twin SOHC 773 cc aircooled. Jantung mekanis tersebut sanggup menyemburkan tenaga 47 Tk dengan torsi 60 Nm. Kesan klasik juga terlihat dari kombinasi yang seimbang antara chasis dengan roda berdiameter besar. Knalpot dengan desain yang sederhana namun memberi kesan elegan.  Kawasaki W800 diciptakan sebagai bentuk penghormatan kepada Kawasaki legendaris W1, brand yang terbentuk selama 45 tahun. Motor ini meluncur di Indonesia dengan harga Rp249.000.000,-, dengan pilihan warna Metallic Ocean Blue dan Ebony (Special Edition). Sebagai informasi, W800 diimpor langsung dari Jepang.  Deputi Departemen Head Sales & Promotion Division KMI Michael Chandra Tanadhi menyebut, model SE paling banyak yakni 40 unit. "Berkaca pada Estrella yang paling banyak dicari model SE. Mereka lebih suka yang full black, garnish-nya juga hitam semua," jelasnya.  KMI mengeluarkan varian Kawasaki W800 di Indonesia untuk melengkapi tipe klasik Kawasaki dan memenuhi kepuasan pecinta motor klasik di Indonesia.  Selain itu, KMI juga memajang 5 motor andalan di booth Kawasaki. Pembelian motor dan apparel Kawasaki bisa dilakukan di booth Kawasaki yang terletak di Hall 1, nomor 1AH, selama GIIAS berlangsung (11 – 21 Agustus 2016). |
| D12 | Daftar mobil baru berbanderol di bawah Rp 200 juta masuk dalam tiga artikel yang paling disorot pembaca setia Liputan6.com.  Di samping itu, motor KTM turun harga dan alasan all new Honda BeAT masuh pakai bohlam tak luput dari perhatian. Berikut top 3 artikel otomotif:   1. Deretan Mobil Baru di Bawah Rp 200 Juta   Pabrikan otomotif sepanjang pekan ini berlomba menampilkan line up andalan mereka di Gaikindo Indonesia International Auto Show (GIIAS) 2016. Beberapa di antara mobil yang dipamerkan pada pameran akbar yang berlokasi di ICE BSD tersebut ada yang model baru, atau meluncur tahun ini.  Bagi sebagian masyarakat Indonesia, pada umumnya mencari mobil dengan harga Rp 200 juta ke bawah. Mobil pada rentang harga ini masih terjangkau dari |

**Tabel 3.7** Daftar Data Latih (Lanjutan)

|  |  |
| --- | --- |
| **Dok** | **Isi Dokumen** |
|  | sisi daya beli serta memiliki fitur yang fungsional untuk pemakaian sehari-hari.  2. Harga Motor KTM Resmi Turun 40 Persen, Mau?  KTM, pabrikan otomotif asal Austria, memastikan mulai hari ini (15/8/2016) akan menjual empat produk mereka lebih murah hingga 40 persen. Hal ini berkaitan dengan rencana pembangunan pabrik perakitan di Gresik, Jawa Timur.  "KTM Duke dan RC sudah bisa diperoleh dengan harga khusus yang sama dengan harga saat nanti kita CKD," ujar Kristianto Goenadi, Presiden Direktur PT Penta Jaya Laju Motor, distributor KTM untuk Indonesia, di booth KTM di hall 1 ICE, BSD City, Tangerang Selatan, Senin (15/8/2016).  3. All New BeAT Masih Pakai Bohlam, Ini Penjelasan Honda  Honda merombak total desain pada All New Honda BeAT eSP. Penampilan bergaya sporty dan futuristik ini sayangnya belum semuanya diimbangi dengan fitur canggih.  Salah satu fitur 'jadul' yang masih jadi andalan pada generasi ketiga BeAT yaitu headlamp yang masih memakai lampu pijar alias bohlam. Padahal, skutik terbaru Honda lainnya telah mengaplikasikan lampu LED. |
| D13 | Honda merombak total desain pada All New Honda BeAT eSP. Penampilan bergaya sporty dan futuristik ini sayangnya belum semuanya diimbangi dengan fitur canggih.  Salah satu fitur 'jadul' yang masih jadi andalan pada generasi ketiga BeAT yaitu headlamp yang masih memakai lampu pijar alias bohlam. Padahal, skutik terbaru Honda lainnya telah mengaplikasikan lampu LED.  "Untuk LED headlight menurut saya fitur yang menambah keuntungan yang besar. Pada saat mengembangkan BeAT baru ini kami sudah mempertimbangkan untuk memakai LED dan dalam tahap analisis kami menemukan kalau penukaran ke lampu LED biaya maintenance cukup tinggi," ungkap Presiden Direktur PT Astra Honda Motor Toshiyuki Inuma saat peluncuran all new Honda BeAT belum lama ini.  Ia menjelaskan bila pemakai BeAT tersebar bahkan hingga seluruh pelosok sehingga jauh dari jangkauan jaringan bengkel resmi. Untuk memudahkan perawatan maka Honda sengaja menunda pemakaian LED sebagai sistem pencahayaan pada BeAT.  "BeAT yang kami produksi digunakan dalam area yang cukup luas, bahkan di daerah-daerah yang tidak terjangkau service network kami. Oleh karena itu mempertimbangkan ketersediaan part, struktur bengkel, maka kali ini kami menunda dulu memakai headlamp LED," tutur pria asal Jepang itu.  Pemakaian lampu hemat listrik tersebut sebenarnya telah menjadi tren di jagat otomotif baik roda empat maupun roda dua. Inuma berjanji, akan memperbanyak model dengan lampu LED.  "Memang jamannya sekarang mengarah ke lampu LED. Kami akan mempertimbangkan memakai LED untuk model-model yang kami kembangkan," pungkasnya. |
| D14 | Di beberapa daerah, dukun beranak seolah masih jadi 'andalan' untuk membantu proses persalinan. Padahal tenaga kesehatan kini sudah banyak yang diterjunkan ke daerah-daerah. Terkait ini, menurut Menkes Dr dr Nila Farid Moelok, Sp.M(K), itu menunjukkan peran dukun beranak yang masih dipercaya |

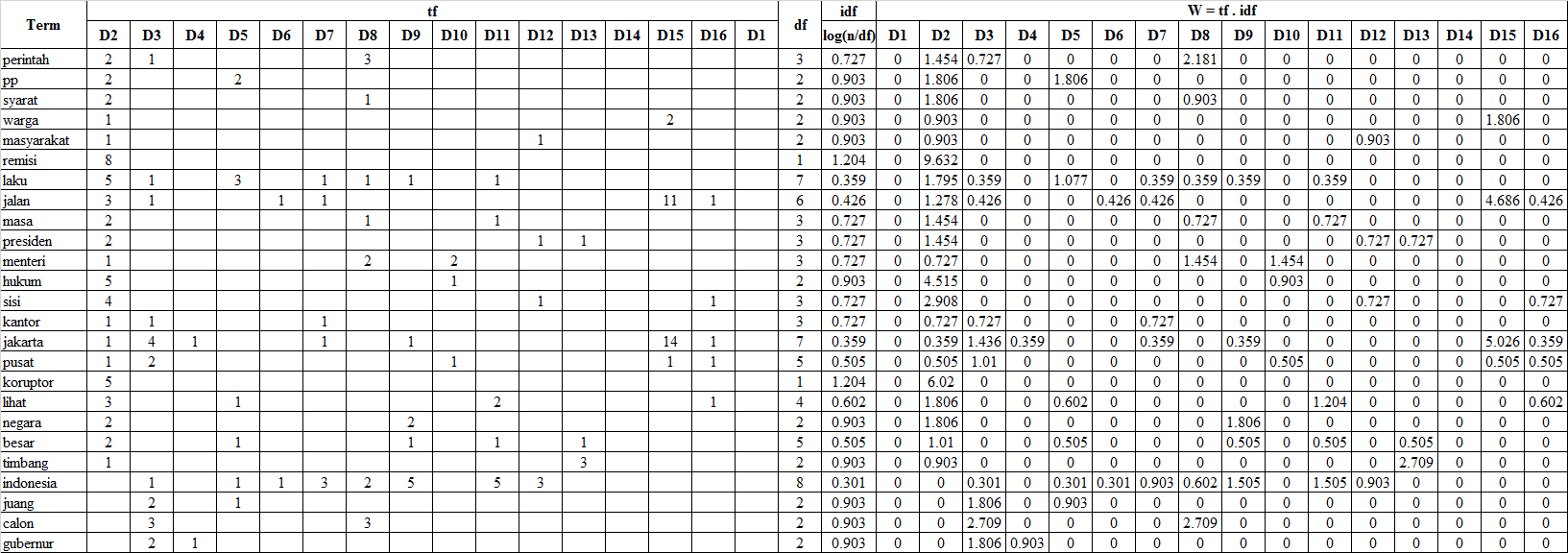
**Tabel 3.7** Daftar Data Latih (Lanjutan)

|  |  |
| --- | --- |
| **Dok** | **Isi Dokumen** |
|  | warga.  Yang dikhawatirkan sebenarnya adalah apabila ibu yang melahirkan mengalami masalah komplikasi sehingga membutuhkan bantuan medis di rumah sakit. Jika dukun beranak tidak peka terhadap masalah ini dan berusaha menanganinya sendiri, dikhawatirkan terjadi hal yang tidak diinginkan.  Mengingat kepercayaan warga di beberapa daerah yang masih besar terhadap sosok dukun beranak, menurut Menkes Nila, tidak bijak jika sosok dukun serta-merta dipinggirkan. Menurutnya, dukun beranak tetap bisa menemani proses persalinan warga.  "Dukun tetap mendampingi di rumah singgahnya, kemudian persalinan dilakukan oleh bidan," saran Menkes dalam penganugerahan tenaga kesehatan teladan tingkat nasional tahun 2016 di Hotel Grand Sahid Jaya, Jalan Jenderal Sudirman, Senin (15/8/2016).  Nila menambahkan baik dukun beranak maupun bidan memiliki kelebihan masing-masing. Untuk dukun beranak memiliki kelebihan dekat di hati masyarakat. Sementara bidan memiliki keterampilan namun kurang memiliki kedekatan hati seperti halnya dukun beranak.  dr Riskiyana Sukandhi Putra, MKes saat masih menjabat Kasub Direktorat Bina Kesehatan Keluarga Berencana Kemenkes, mengatakan peran dukun beranak untuk memberi pertolongan non-medis bisa diwujudkan melalui rumah tunggu. Rumah tunggu merupakan rumah singgah bagi ibu-ibu yang ingin melahirkan. Di rumah tunggu itulah para ibu akan diajarkan hal-hal yang bersifat non-medis.  "Nanti ibu-ibu ini akan diajarkan oleh dukun bayi ini bagaimana cara memandikan anak, merawat tali pusat, menyusui, terus nanti dikasih tahu anak harus diimunisasi kapan, kalau anak mulai panas nanti diginiin ya bu. Bayangkan dengan tingkat pendidikan yang tidak terlalu tinggi, pengetahuan seperti itu kan pasti berguna sekali," tutur dr Riski.  Lokasi rumah tunggu biasanya berada di dekat puskesmas atau rumah sakit. Sehingga, sambil menunggu persalinan, para ibu bisa belajar bagaimana merawat anaknya. "Terutama yang di daerah, bisa dua tiga atau lima hari sebelum bersalin sudah di sana. Kalau seperti di Rote, untuk ke fasilitas kesehatan butuh waktu lima hari, maka dia harus di rumah singgah seminggu sebelum melahirkan kan," lanjut Riski. |
| D15 | Hujan deras yang mengguyur sebagian wilayah ibu kota menyebabkan genangan hingga banjir di beberapa titik.  Berdasarkan twitter @BPBDJakarta, genangan terjadi di delapan titik.  1. Pukul 21.33 WIB, Jalan Bahari RT 10/07, Kelurahan Gandaria Selatan, tergenang 10-40 cm  2. Pukul 21.36 WIB, Jalan Rasuna Said, Jakarta Selatan tergenang 10-30 cm  3. Pukul 21.38 WIB, Jalan Letjen S Parman depan Untar Kelurahan Tomang, Kecamatan Grogol Petamburan, Jakarta Barat tergenang 10-20 cm  4. Pukul 21.42 WIB, Jalan H. Montong, Kecamatan Ciganjur tergenang 10-30 cm;  5. Pukul 21.51 WIB, Jalan Kedoya Raya, Jakarta Barat tergenang 10-30 cm;  6. Pukul 21.53 WIB, Jalan Boulevard Raya, Kelapa Gading, Jakarta Utara, tergenang 10-30 cm  7. Pukul 22.08 WIB, Jalan Kapuk Raya Kelurahan Kapuk, Jakarta Barat tergenang 20-30 cm  8. Pukul 22.12 WIB, Jalan Karet Pasar Baru Timur, Kelurahan Karet Tengsin, Kecamatan Tanah Abang, Jakarta Pusat tergenang 20-30 cm.  Banjir  1. Pukul 20.38 WIB, Jalan Jagakarsa 2 Gang H Masmun belakang TPU, Jakarta Selatan, terendam banjir setinggi 40 sentimeter. Sebanyak 30 rumah warga terendam |

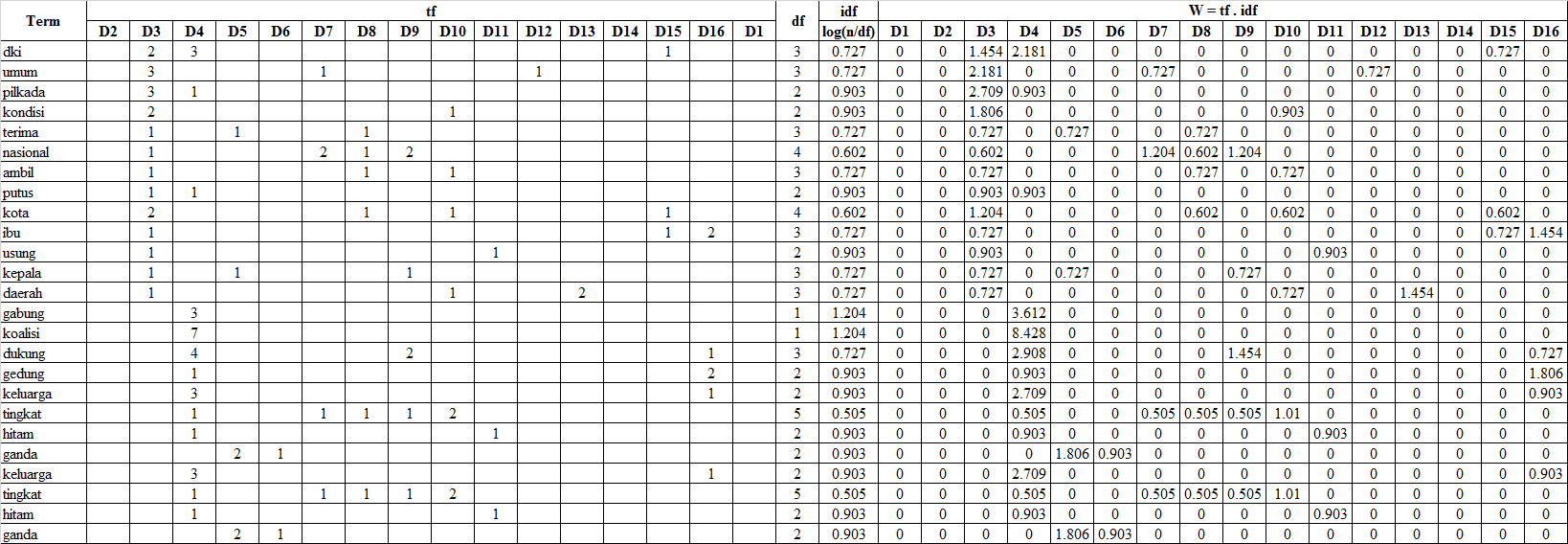
**Tabel 3.7** Daftar Data Latih (Lanjutan)

|  |  |
| --- | --- |
| **Dok** | **Isi Dokumen** |
|  | 2. Pukul 21.13 WIB, banjir setinggi 50-100 sentimeter merendam RW 03, Kelurahan Wijaya Kesuma, Kecamatan Grogol Petamburan, Jakarta Barat;  3. Pukul 21.42 WIB, banjir setinggi 50-120 sentimeter merendam RT 13, 12, 11, 8, 5, dan 2 RW di Kelurahan Ragunan, Kecamatan Pasar Minggu, Jakarta Selatan;  4. Pukul 21.49 WIB, banjir setinggi 10-70 sentimeter merendam RT 06/07, Pondok Labu, Jakarta Selatan;  5. Pukul 21.57 WIB, banjir merendam RW 09 Kelurahan Kramat Pela, Kecamatan Kebayoran Baru, Jakarta Selatan.  Pohon tumbang  Pukul 20.45 WIB, sebuah pohon mahoni sempal tumbang di Jalan Arteri Permata Hijau dekat jembatan arah Kebayoran Lama, Jakarta Selatan. Kini, pohon tumbang tersebut tengah ditangani oleh dua unit petugas Dinas Pertamanan DKI Jakarta serta PPSU.  Selain pohon tumbang, longsor juga terjadi di Jalan Melati 2, RT 09, RW 02, Kelurahan Pondok Labu, Jakarta Selatan, pukul 20.00 WIB. Saat ini, tengah dilakukan penyedotan dan mengevakuasi 75 jiwa warga terdampak. |
| D16 | Dengan bermain dalam film baru Pinky Promise, artis peran, model, dan pembawa acara Maudy Koesnaedi (41) mendapat pelajaran tentang hidup dengan anak yang menderita kanker.  Dalam film itu, Maudy berperan sebagai seorang ibu yang mengalami dilema dalam mendampingi putrinya yang didiagnosis terkena kanker payudara.  Dalam film tersebut pula, suami Maudy berbeda pendapat dengan anak mereka mengenai pengobatan yang harus dijalani oleh sang anak.  "Saya punya dilema untuk men-support anak saya yang terkena kanker payudara," kata Maudy mengenai perannya dalam film itu, dalam jumpa pers peluncuran trailer dan poster film tersebut di Plaza Senayan XXI, Jakarta Pusat, Jumat (19/8/2016).  "Saya tahu betul mendampingi anak saya. Di luar itu, saya sebagai istri harus menghormati suami yang memiliki pendapat lain," sambungnya.  Film karya Guntur Soeharjanto itu menggambarkan bagaimana orang-orang terdekat memberi dukungan bagi penderita kanker.  "Film ini lebih kepada keluarga menempatkan diri menghadapi survivor (penyintas) kanker payudara. Gimana kadang kan penginnya ngebantu dan perhatian, tapi jadinya ngeganggu," ujar ibu satu anak laki-laki ini.  "Survivor kan punya motivasi dan keyakinan sendiri dalam masalahnya dan kita juga ingin membantunya. Nah, ini buat pembelajaran ke saya, harus lihat dari sisi mananya," ujarnya lagi.  Pinky Promise dibintangi juga oleh Agni Pratista, Dhea Seto, Ira Maya Sopha, Dea Ananda, Chelsea Islan, Alexandra Gottardo, Gunawan, Maudy Koesnaedi, dan Ringgo Agus Rachman.  Film itu akan mulai diputar di gedung-gedung bioskop pada 13 Oktober 2016. |

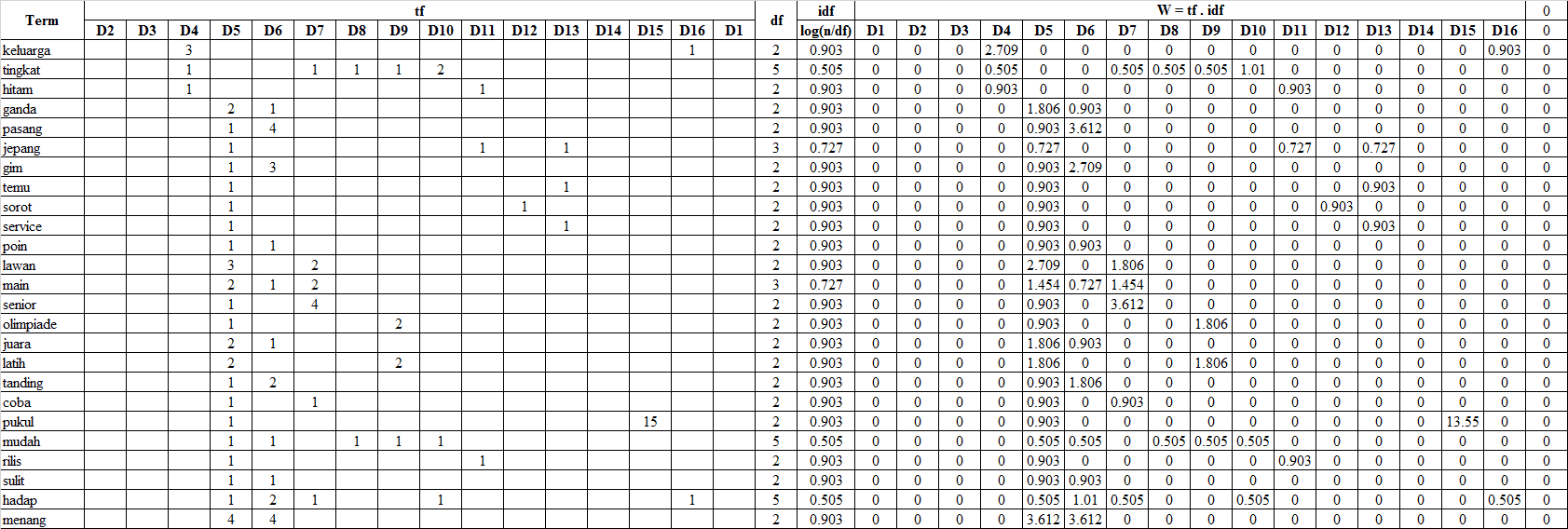
**Tabel 3.8** Pembobotsan Tf-Idf



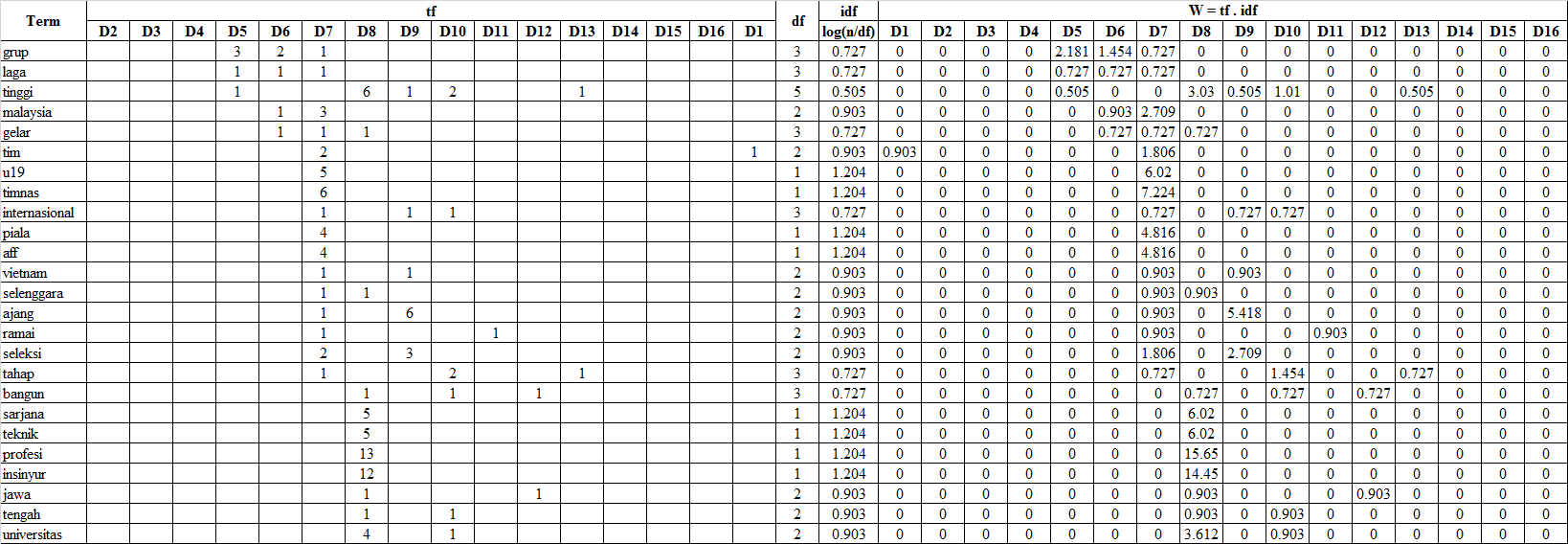
**Tabel 3.8** Pembobotsan Tf-Idf (Lanjutan)



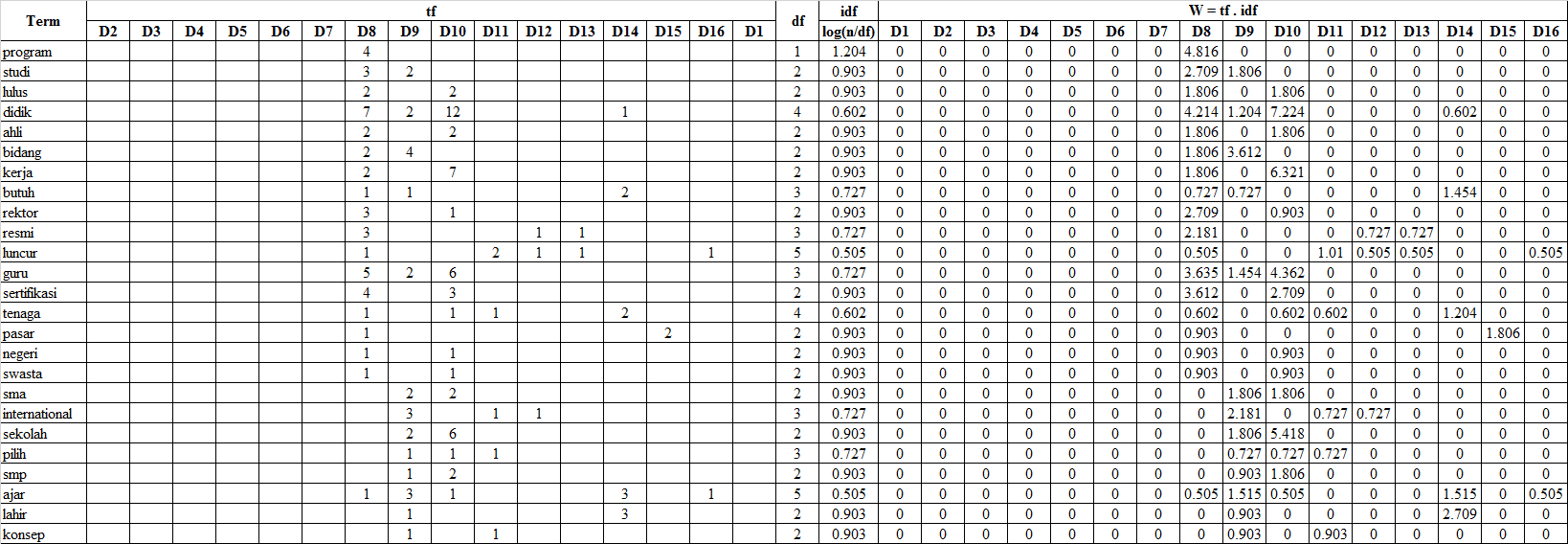
**Tabel 3.8** Pembobotsan Tf-Idf (Lanjutan)



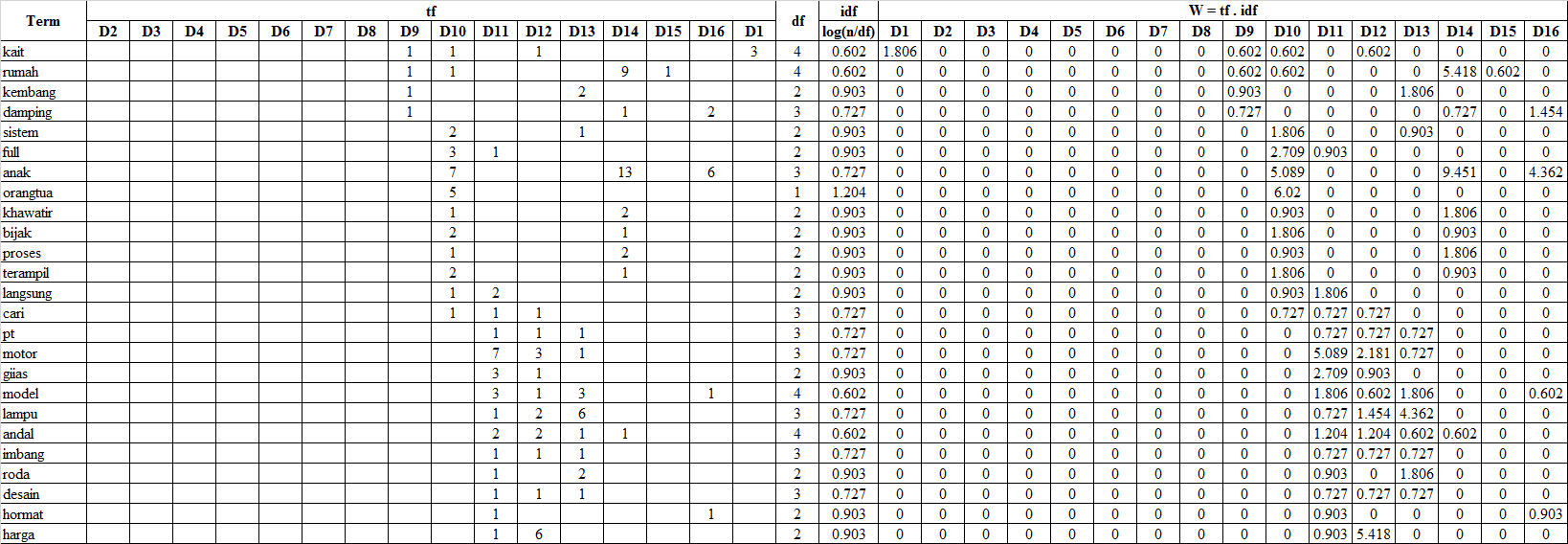
**Tabel 3.8** Pembobotsan Tf-Idf (Lanjutan)



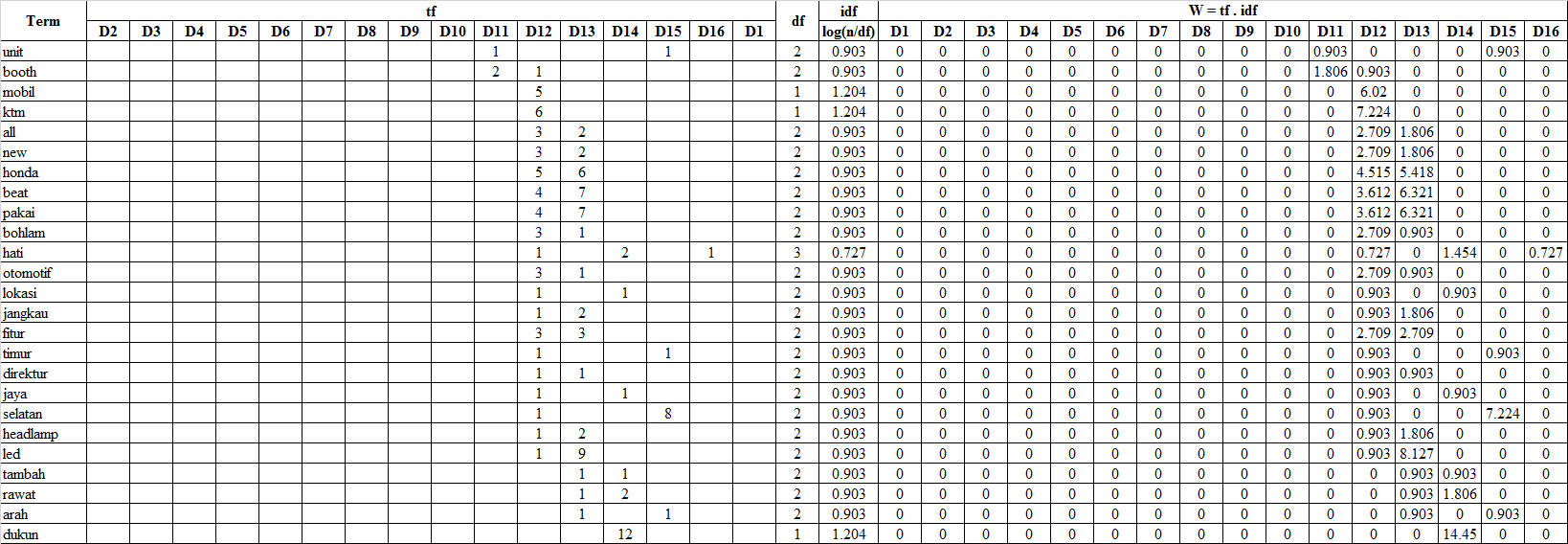
**Tabel 3.8** Pembobotsan Tf-Idf (Lanjutan)



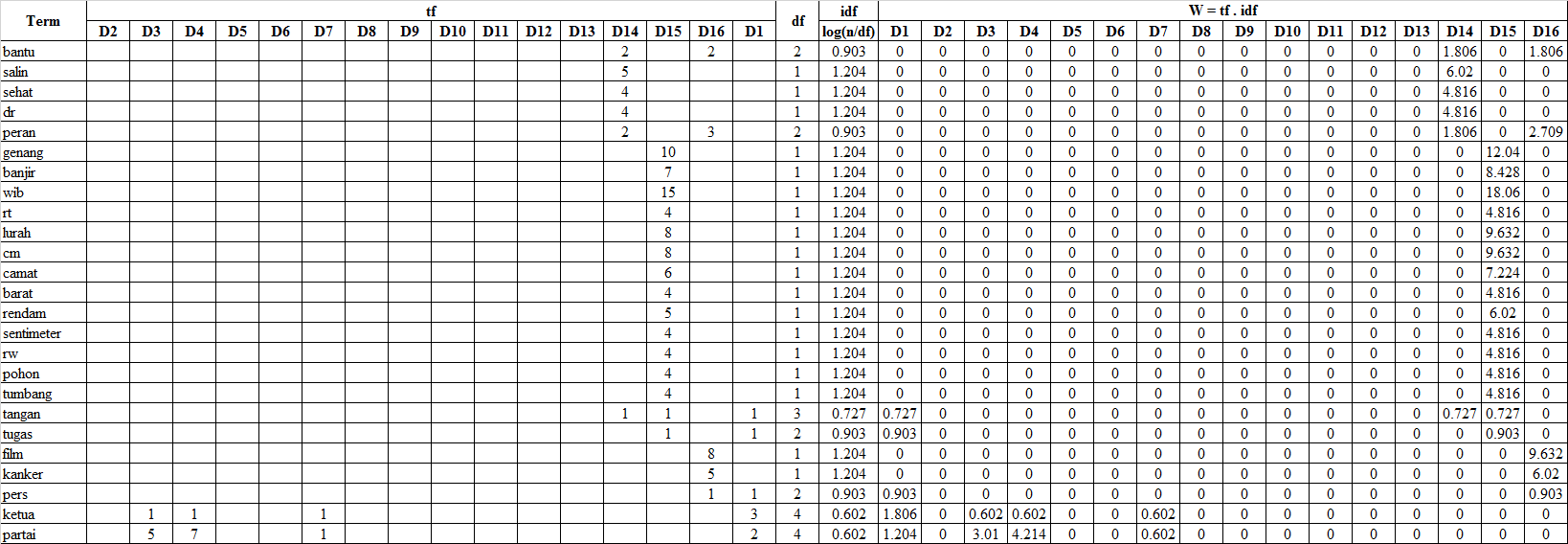
**Tabel 3.8** Pembobotsan Tf-Idf (Lanjutan)



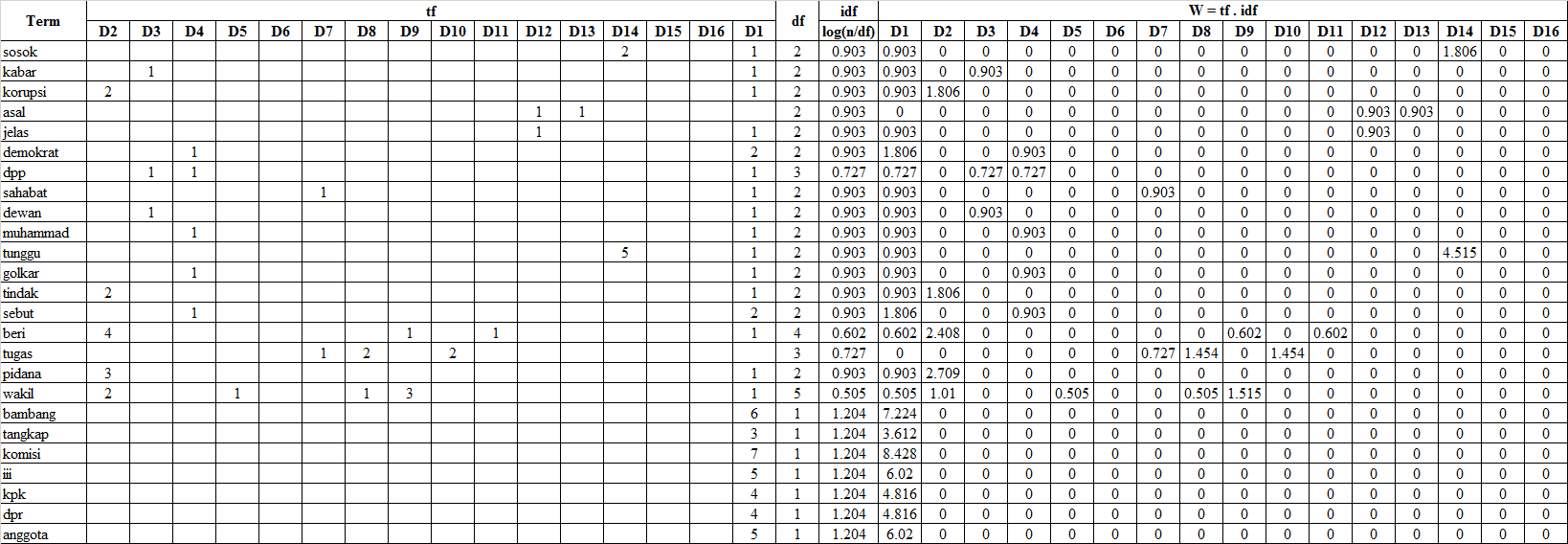
**Tabel 3.8** Pembobotsan Tf-Idf (Lanjutan)



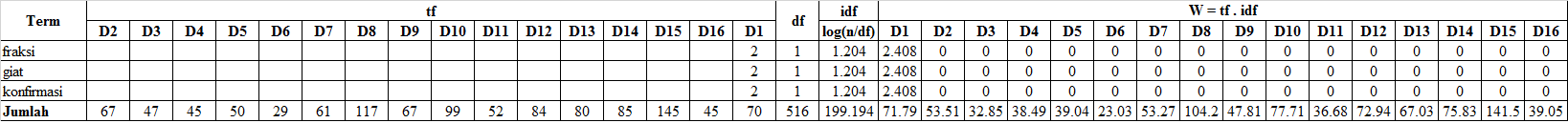
**Tabel 3.8** Pembobotsan Tf-Idf (Lanjutan)



**Tabel 3.8** Pembobotsan Tf-Idf (Lanjutan)



**Tabel 3.8** Pembobotsan Tf-Idf (Lanjutan)



1. **Pengkategorian dengan *K-Nearest Neighbour***

Proses pengkategorian metode *k-nearest neighbour* menggunakan rumus *cosine similarity* untuk menghitung kemiripan jarak titik uji dan jarak titik latih. Langkah yang dilakukan untuk menghitung kemiripannya dilakukan sebagai berikut:

Langkah 1: Setelah mendapatkan hasil pembobotan tf-idf hitung hasil perkalian skalar antara dokumen data uji dengan semua dokumen data latih. Hasil dari perkalian tersebut dijumlahkan sesuai dengan dokumen masing-masing.

Langkah 2: Hitung panjang setiap dokumen, termasuk data uji. Caranya dengan mengkuadratkan bobot setiap *term*  dalam setiap dokumen, kemudian jumlahkan nilai kuadratnya dan kuadratkan hasil penjumlahannya.

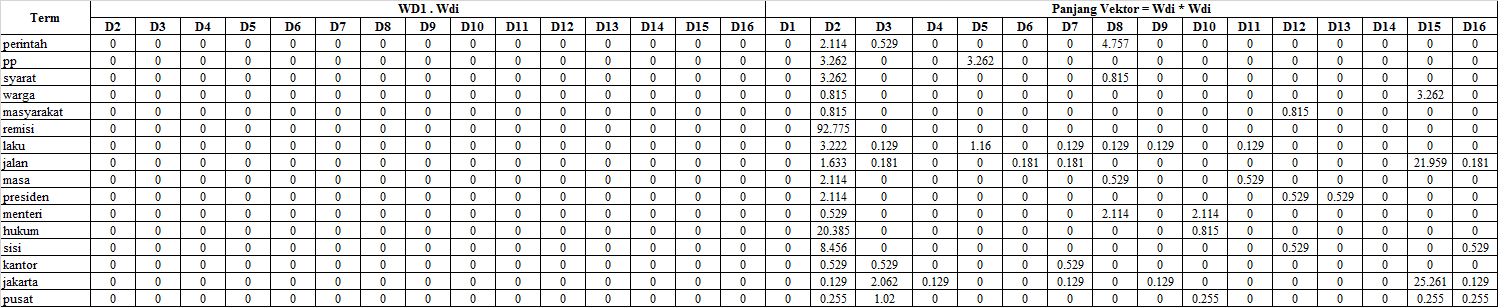
Langkah 3: Terapkan rumus *cosine similarity.* Hitung kemiripan data uji dengan setiap data latih.

Langkah 4: Urutkan hasil perhitungan kemiripan.

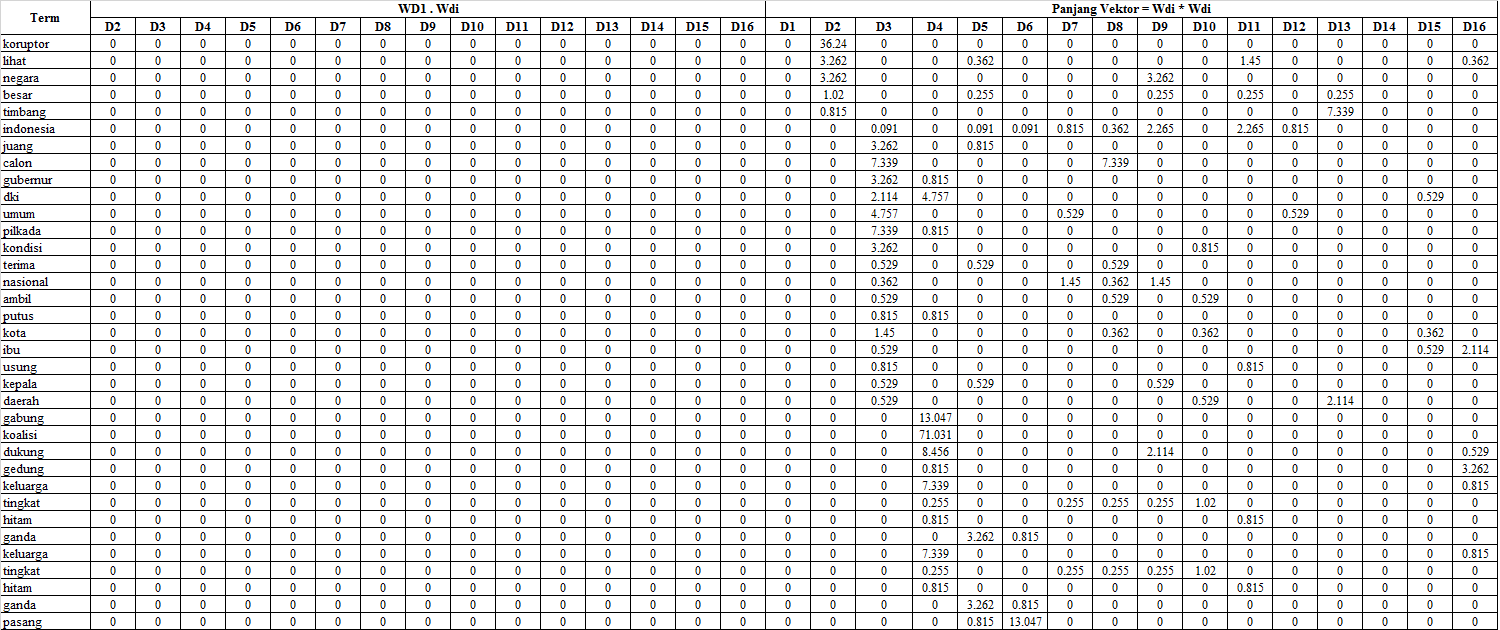
Langkah 5: Ambil sebanyak k (k=5) yang paling tinggi nilai kemiripannya dengan data uji. Tentukan kelas dari data uji.

Untuk melihat hasil dari perhitungannya dapat dilihat pada Tabel 3.9 di bawah ini.

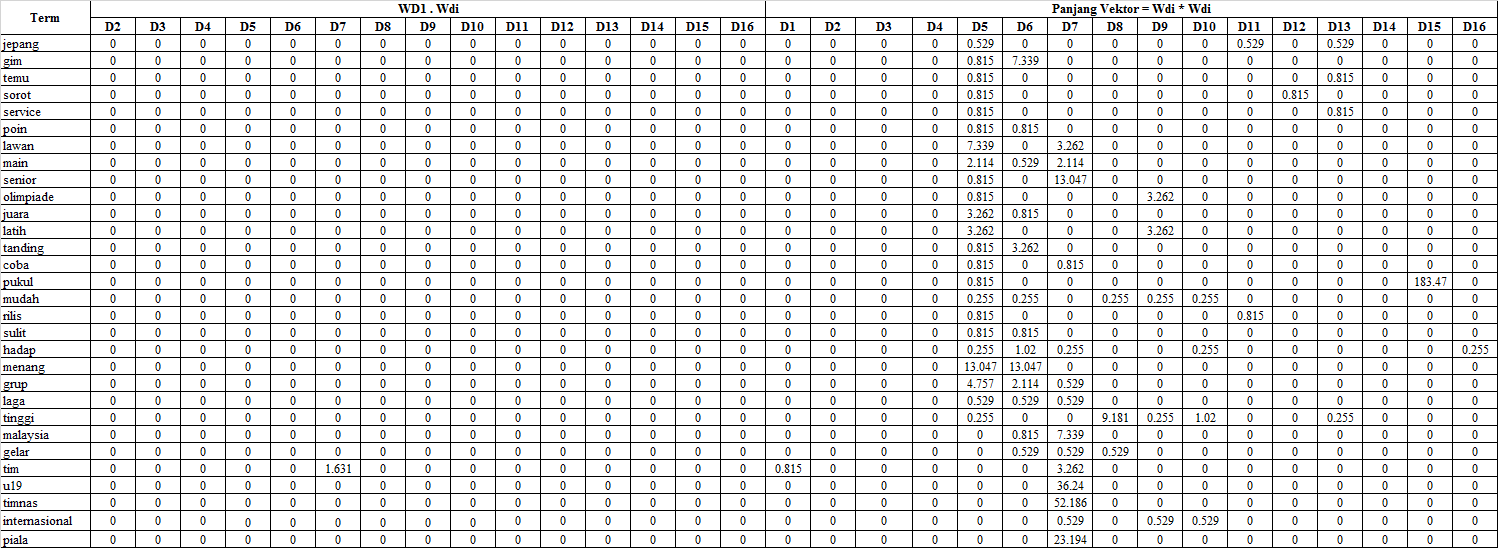
**Tabel 3.9** *Cosine Similarity*



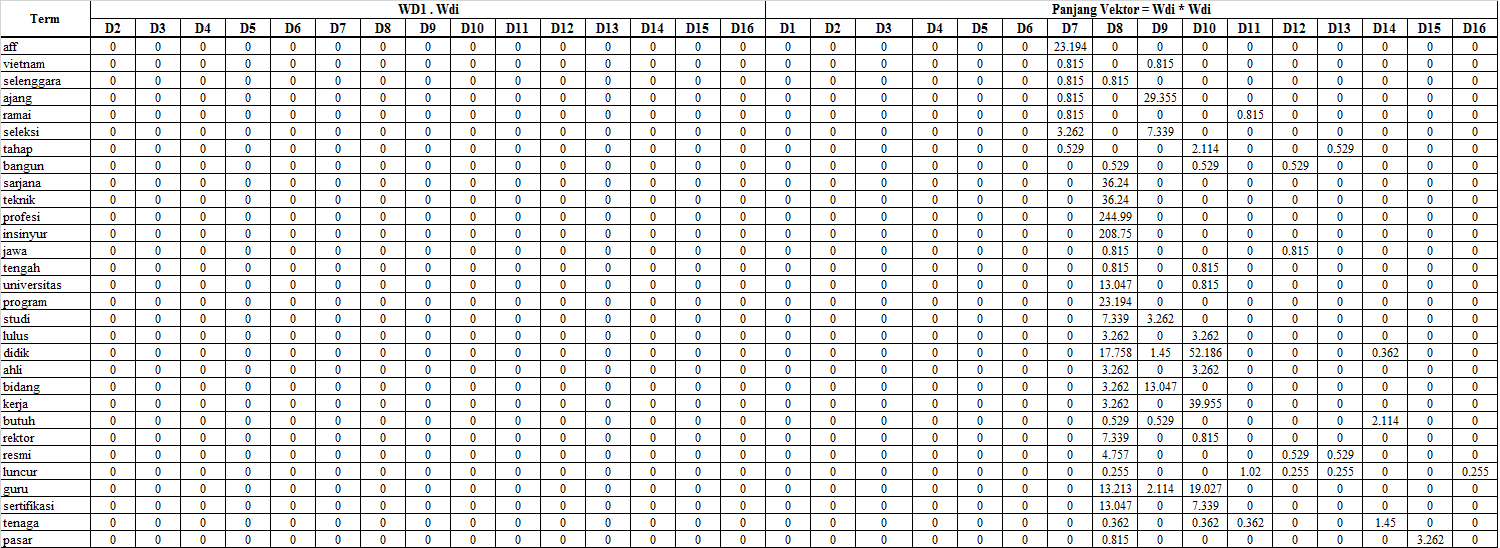
**Tabel 3.9** *Cosine Similarity* (Lanjutan)



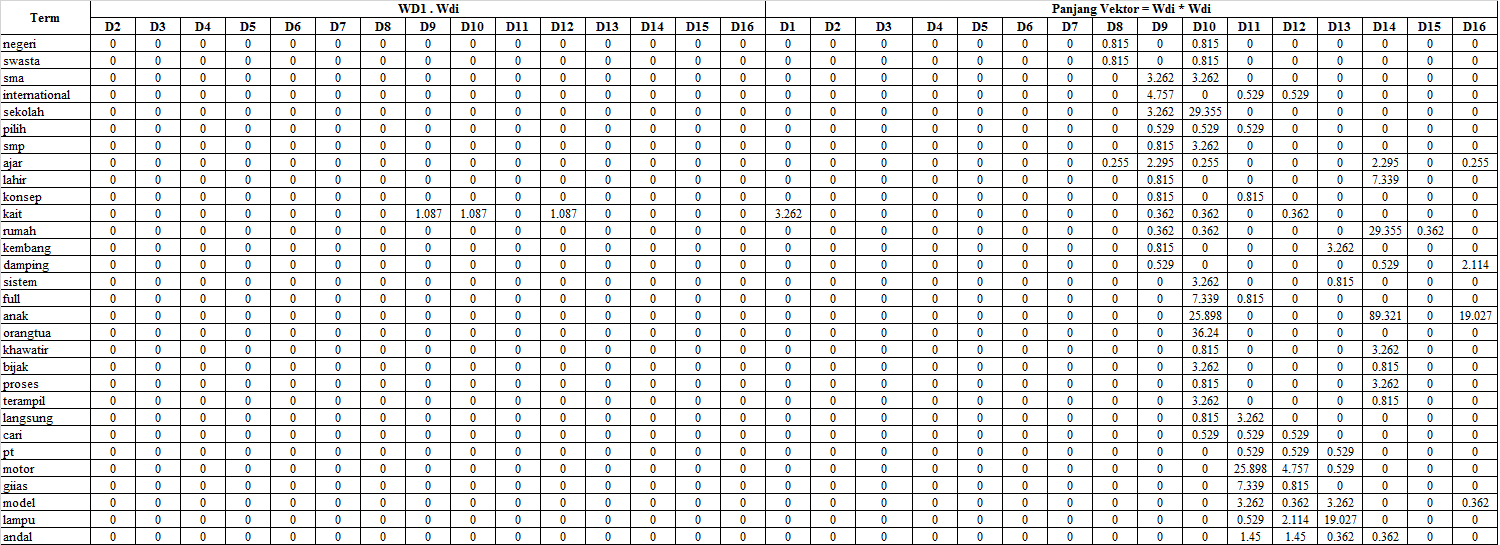
**Tabel 3.9** *Cosine Similarity* (Lanjutan)



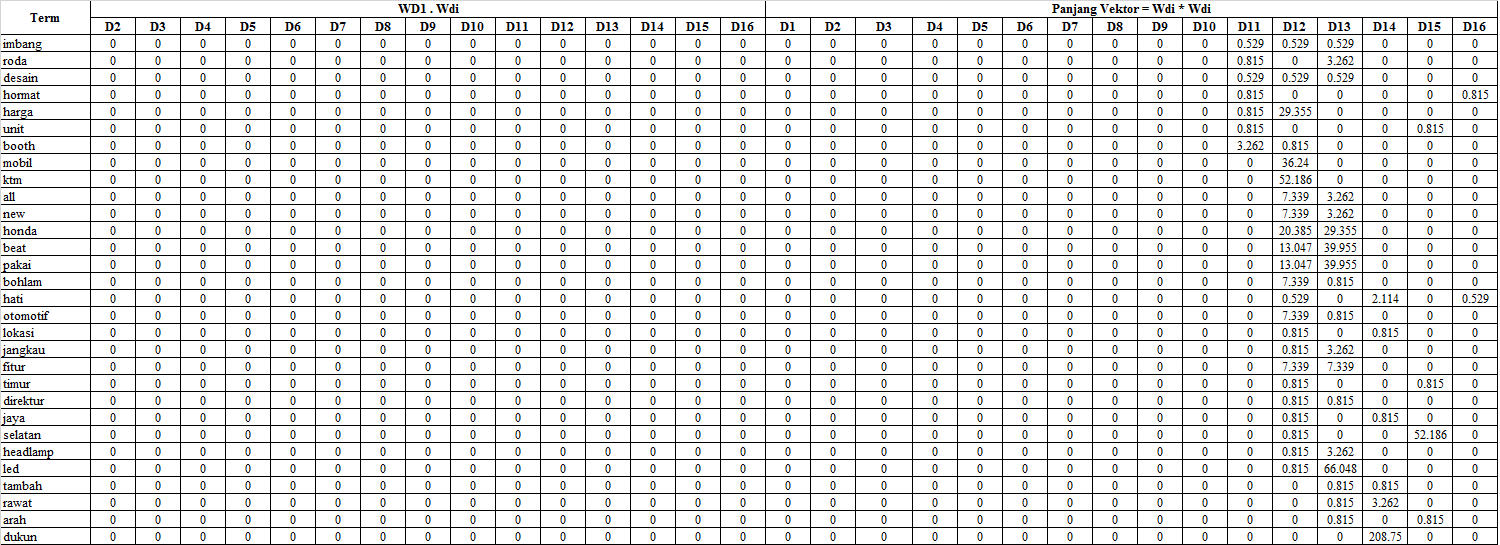
**Tabel 3.9** *Cosine Similarity* (Lanjutan)



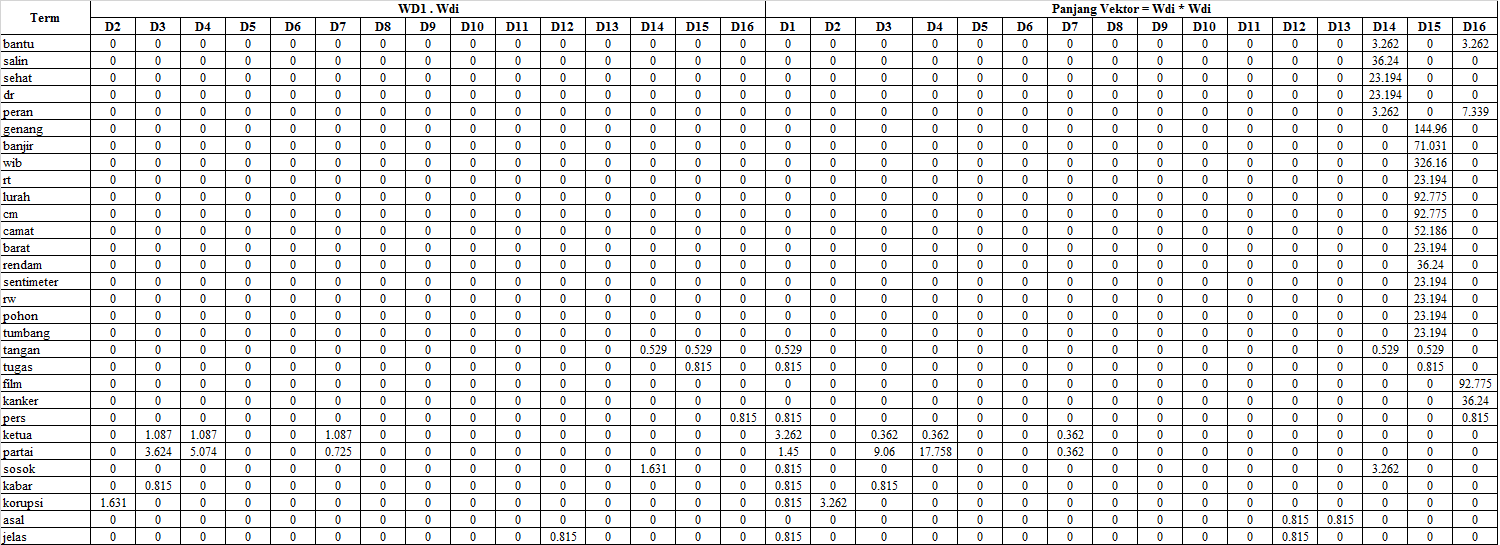
**Tabel 3.9** *Cosine Similarity* (Lanjutan)



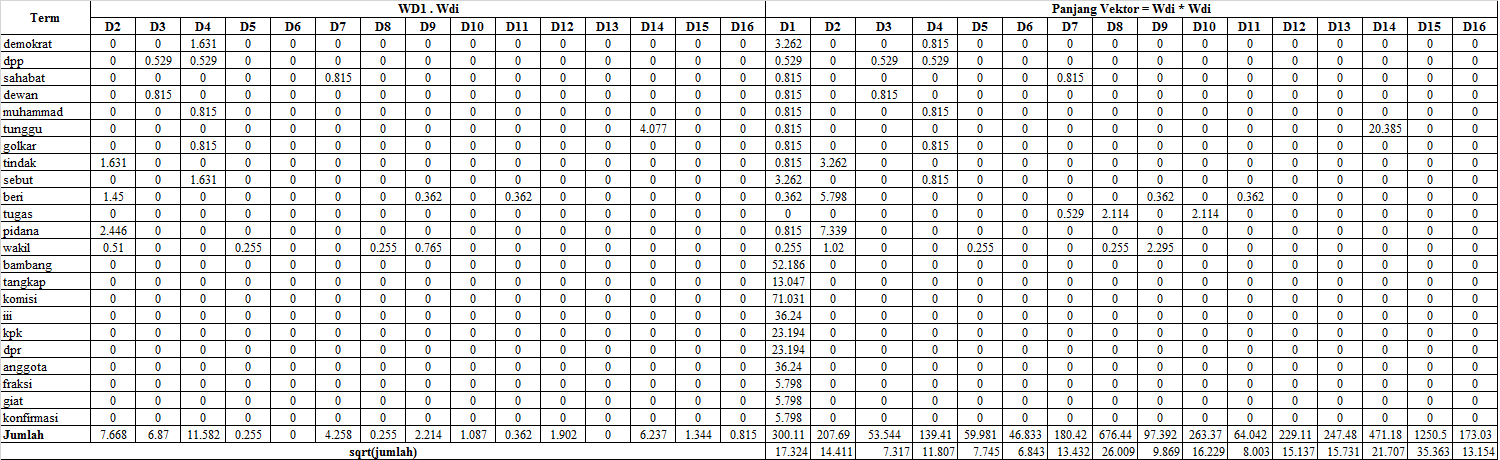
**Tabel 3.9** *Cosine Similarity* (Lanjutan)



**Tabel 3.9** *Cosine Similarity* (Lanjutan)



**Tabel 3.9** *Cosine Similarity* (Lanjutan)



Setelah mendapatkan panjang vector tiap dokumen, langkah selanjutnya yaitu menerapkan rumus *cosine similarity* untuk menghitung kemiripan data uji (D1) dengan data latih.

Cos (D1, Dt) = ∑wd2 / sqrt (∑wd12 ) \* sqrt (∑wdt2)

Cos (D1, D2) = 7.67 / (17.325 \* 14.411) = 0.031

Cos (D1, D3) = 6.88 / (17.325 \* 7.318) = 0.054

Cos (D1, D4) = 11.59 / (17.325 \* 11.817) = 0.057

Cos (D1, D5) = 0.26 / (17.325 \* 7.7504) = 0.002

Cos (D1, D6) = 0 / (17.325 \* 6.8461) = 0

Cos (D1, D7) = 4.24 / (17.325 \* 13.435) = 0.018

Cos (D1, D8) = 0.26 / (17.325 \* 26.011) = 0.001

Cos (D1, D9) = 2.22 / (17.325 \* 9.874) = 0.013

Cos (D1, D10) = 1.09 / (17.325 \* 16.23) = 0.004

Cos (D1, D11) = 0.36 / (17.325 \* 8.007) = 0.003

Cos (D1, D12) = 1.91 / (17.325 \* 15.141) = 0.007

Cos (D1, D13) = 0 / (17.325 \* 15.737) = 0

Cos (D1, D14) = 6.24 / (17.325 \* 21.708) = 0.0167

Cos (D1, D15) = 1.35 / (17.325 \* 35.3367) = 0.002

Cos (D1, D16) = 0.82 / (17.325 \* 13.155) = 0.004

Langkah selanjutnya adalah mengurutkan hasil perhitungan *cosine similarity* dari nilai terbesar ke nilai lebih kecil, yang ditunjukan pada Tabel 3.10.

**Tabel 3.10** Hasil Pengurutan Nilai *Cosine Similarity*

| **No. Urut** | **Dokumen** | **Nilai** |
| --- | --- | --- |
| 1 | D4 | 0.057 |
| 2 | D3 | 0.054 |
| 3 | D2 | 0.031 |
| 4 | D7 | 0.018 |
| 5 | D14 | 0.017 |
| 6 | D9 | 0.013 |
| 7 | D12 | 0.007 |
| 8 | D10 | 0.004 |
| 9 | D16 | 0.004 |
| 10 | D11 | 0.003 |
| 11 | D15 | 0.002 |
| 12 | D5 | 0.002 |
| 13 | D8 | 0.001 |
| 14 | D6 | 0 |
| 15 | D13 | 0 |

Dari tabel di atas ambil k sebanyak 5 dokumen dengan nilai teratas, yaitu D4, D3, D2, D7, dan D14. Untuk kategori Politik diwakili D4, D3, dan D2, kategori Olahraga diwakili oleh D7, sedangkan kategori Umum diwakili oleh D14. Kelas yang paling banyak kemunculannya yaitu kelas dengan kategori Politik. Jadi, D1 termasuk ke dalam kategori **Politik**.

1. **Pengkategorian dengan *Naïve Bayes***

Proses kategori dengan *naïve bayes* dilakukan dengan dua tahap, yaitu proses pelatihan dan klasifikasi. Tahap klasifikasi dilakukan dengan memilih *term* yang mungkin muncul dalam dokumen sampel yang menjadi representasi dokumen. Tahap klasifikasi dilakukan dengan menentukan nilai kategori dari dokumen berdasarkan *term* yang muncul dalam dokumen yang diklasifikasikan. Tabel 3.11 menunjukkan hasil perhitungan *naïve bayes.*

**Tabel 3.11** Hasil Hitung *Naïve Bayes*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Term** | **PVj** | P(Wk | Vj) | | | | |
| Politik | Olahraga | Pendidikan | Otomotif | Umum |
| perintah | 0.6 | 0.017 | 0.005 | 0.011 | 0.003 | 0.003 |
| pp | 0.6 | 0.013 | 0.014 | 0.003 | 0.003 | 0.003 |
| syarat | 0.6 | 0.013 | 0.005 | 0.006 | 0.003 | 0.003 |
| warga | 0.6 | 0.009 | 0.005 | 0.003 | 0.003 | 0.009 |
| masyarakat | 0.6 | 0.009 | 0.005 | 0.003 | 0.007 | 0.003 |
| remisi | 0.6 | 0.039 | 0.005 | 0.003 | 0.003 | 0.003 |
| laku | 0.6 | 0.031 | 0.024 | 0.008 | 0.007 | 0.003 |
| jalan | 0.6 | 0.022 | 0.014 | 0.003 | 0.003 | 0.038 |
| masa | 0.6 | 0.013 | 0.005 | 0.006 | 0.007 | 0.003 |
| presiden | 0.6 | 0.013 | 0.005 | 0.003 | 0.01 | 0.003 |
| menteri | 0.6 | 0.009 | 0.005 | 0.014 | 0.003 | 0.003 |
| hukum | 0.6 | 0.026 | 0.005 | 0.006 | 0.003 | 0.003 |
| sisi | 0.6 | 0.022 | 0.005 | 0.003 | 0.007 | 0.006 |
| kantor | 0.6 | 0.013 | 0.01 | 0.003 | 0.003 | 0.003 |
| jakarta | 0.6 | 0.031 | 0.01 | 0.006 | 0.003 | 0.046 |
| pusat | 0.6 | 0.017 | 0.005 | 0.006 | 0.003 | 0.009 |
| koruptor | 0.6 | 0.026 | 0.005 | 0.003 | 0.003 | 0.003 |
| lihat | 0.6 | 0.017 | 0.01 | 0.003 | 0.01 | 0.006 |
| negara | 0.6 | 0.013 | 0.005 | 0.008 | 0.003 | 0.003 |
| besar | 0.6 | 0.013 | 0.01 | 0.006 | 0.01 | 0.003 |
| timbang | 0.6 | 0.009 | 0.005 | 0.003 | 0.014 | 0.003 |
| indonesia | 0.6 | 0.009 | 0.029 | 0.023 | 0.031 | 0.003 |
| juang | 0.6 | 0.013 | 0.01 | 0.003 | 0.003 | 0.003 |
| calon | 0.6 | 0.017 | 0.005 | 0.011 | 0.003 | 0.003 |
| gubernur | 0.6 | 0.017 | 0.005 | 0.003 | 0.003 | 0.003 |
| dki | 0.6 | 0.026 | 0.005 | 0.003 | 0.003 | 0.006 |
| umum | 0.6 | 0.017 | 0.01 | 0.003 | 0.007 | 0.003 |
| pilkada | 0.6 | 0.022 | 0.005 | 0.003 | 0.003 | 0.003 |
| kondisi | 0.6 | 0.013 | 0.005 | 0.006 | 0.003 | 0.003 |
| terima | 0.6 | 0.009 | 0.01 | 0.006 | 0.003 | 0.003 |
| nasional | 0.6 | 0.009 | 0.014 | 0.011 | 0.003 | 0.003 |
| ambil | 0.6 | 0.009 | 0.005 | 0.008 | 0.003 | 0.003 |
| putus | 0.6 | 0.013 | 0.005 | 0.003 | 0.003 | 0.003 |
| kota | 0.6 | 0.013 | 0.005 | 0.008 | 0.003 | 0.006 |
| ibu | 0.6 | 0.009 | 0.005 | 0.003 | 0.003 | 0.012 |
| usung | 0.6 | 0.009 | 0.005 | 0.003 | 0.007 | 0.003 |
| kepala | 0.6 | 0.009 | 0.01 | 0.006 | 0.003 | 0.003 |
| daerah | 0.6 | 0.009 | 0.005 | 0.006 | 0.01 | 0.003 |
| gabung | 0.6 | 0.017 | 0.005 | 0.003 | 0.003 | 0.003 |

**Tabel 3.11** Hasil Hitung *Naïve Bayes* (Lanjutan)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Term** | **PVj** | P(Wk | Vj) | | | | |
| Politik | Olahraga | Pendidikan | Otomotif | Umum |
|  |  |  |  |  |  |  |
| koalisi | 0.6 | 0.035 | 0.005 | 0.003 | 0.003 | 0.003 |
| dukung | 0.6 | 0.022 | 0.005 | 0.008 | 0.003 | 0.006 |
| gedung | 0.6 | 0.009 | 0.005 | 0.003 | 0.003 | 0.009 |
| keluarga | 0.6 | 0.017 | 0.005 | 0.003 | 0.003 | 0.006 |
| tingkat | 0.6 | 0.009 | 0.01 | 0.014 | 0.003 | 0.003 |
| hitam | 0.6 | 0.009 | 0.005 | 0.003 | 0.007 | 0.003 |
| ganda | 0.6 | 0.004 | 0.019 | 0.003 | 0.003 | 0.003 |
| keluarga | 0.6 | 0.017 | 0.005 | 0.003 | 0.003 | 0.006 |
| tingkat | 0.6 | 0.009 | 0.01 | 0.014 | 0.003 | 0.003 |
| hitam | 0.6 | 0.009 | 0.005 | 0.003 | 0.007 | 0.003 |
| ganda | 0.6 | 0.004 | 0.019 | 0.003 | 0.003 | 0.003 |
| pasang | 0.6 | 0.004 | 0.029 | 0.003 | 0.003 | 0.003 |
| jepang | 0.6 | 0.004 | 0.01 | 0.003 | 0.01 | 0.003 |
| gim | 0.6 | 0.004 | 0.024 | 0.003 | 0.003 | 0.003 |
| temu | 0.6 | 0.004 | 0.01 | 0.003 | 0.007 | 0.003 |
| sorot | 0.6 | 0.004 | 0.01 | 0.003 | 0.007 | 0.003 |
| service | 0.6 | 0.004 | 0.01 | 0.003 | 0.007 | 0.003 |
| poin | 0.6 | 0.004 | 0.014 | 0.003 | 0.003 | 0.003 |
| lawan | 0.6 | 0.004 | 0.029 | 0.003 | 0.003 | 0.003 |
| main | 0.6 | 0.004 | 0.029 | 0.003 | 0.003 | 0.003 |
| senior | 0.6 | 0.004 | 0.029 | 0.003 | 0.003 | 0.003 |
| olimpiade | 0.6 | 0.004 | 0.01 | 0.008 | 0.003 | 0.003 |
| juara | 0.6 | 0.004 | 0.019 | 0.003 | 0.003 | 0.003 |
| latih | 0.6 | 0.004 | 0.014 | 0.008 | 0.003 | 0.003 |
| tanding | 0.6 | 0.004 | 0.019 | 0.003 | 0.003 | 0.003 |
| coba | 0.6 | 0.004 | 0.014 | 0.003 | 0.003 | 0.003 |
| pukul | 0.6 | 0.004 | 0.01 | 0.003 | 0.003 | 0.046 |
| mudah | 0.6 | 0.004 | 0.014 | 0.011 | 0.003 | 0.003 |
| rilis | 0.6 | 0.004 | 0.01 | 0.003 | 0.007 | 0.003 |
| sulit | 0.6 | 0.004 | 0.014 | 0.003 | 0.003 | 0.003 |
| hadap | 0.6 | 0.004 | 0.024 | 0.006 | 0.003 | 0.006 |
| menang | 0.6 | 0.004 | 0.043 | 0.003 | 0.003 | 0.003 |
| grup | 0.6 | 0.004 | 0.033 | 0.003 | 0.003 | 0.003 |
| laga | 0.6 | 0.004 | 0.019 | 0.003 | 0.003 | 0.003 |
| tinggi | 0.6 | 0.004 | 0.01 | 0.028 | 0.007 | 0.003 |
| malaysia | 0.6 | 0.004 | 0.024 | 0.003 | 0.003 | 0.003 |
| gelar | 0.6 | 0.004 | 0.014 | 0.006 | 0.003 | 0.003 |
| tim | 0.6 | 0.004 | 0.014 | 0.003 | 0.003 | 0.003 |

**Tabel 3.11** Hasil Hitung *Naïve Bayes* (Lanjutan)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Term** | **PVj** | P(Wk | Vj) | | | | |
| Politik | Olahraga | Pendidikan | Otomotif | Umum |
| u19 | 0.6 | 0.004 | 0.029 | 0.003 | 0.003 | 0.003 |
| timnas | 0.6 | 0.004 | 0.033 | 0.003 | 0.003 | 0.003 |
| internasional | 0.6 | 0.004 | 0.01 | 0.008 | 0.003 | 0.003 |
| piala | 0.6 | 0.004 | 0.024 | 0.003 | 0.003 | 0.003 |
| aff | 0.6 | 0.004 | 0.024 | 0.003 | 0.003 | 0.003 |
| vietnam | 0.6 | 0.004 | 0.01 | 0.006 | 0.003 | 0.003 |
| selenggara | 0.6 | 0.004 | 0.01 | 0.006 | 0.003 | 0.003 |
| ajang | 0.6 | 0.004 | 0.01 | 0.02 | 0.003 | 0.003 |
| ramai | 0.6 | 0.004 | 0.01 | 0.003 | 0.007 | 0.003 |
| seleksi | 0.6 | 0.004 | 0.014 | 0.011 | 0.003 | 0.003 |
| tahap | 0.6 | 0.004 | 0.01 | 0.008 | 0.007 | 0.003 |
| bangun | 0.6 | 0.004 | 0.005 | 0.008 | 0.007 | 0.003 |
| sarjana | 0.6 | 0.004 | 0.005 | 0.017 | 0.003 | 0.003 |
| teknik | 0.6 | 0.004 | 0.005 | 0.017 | 0.003 | 0.003 |
| profesi | 0.6 | 0.004 | 0.005 | 0.04 | 0.003 | 0.003 |
| insinyur | 0.6 | 0.004 | 0.005 | 0.037 | 0.003 | 0.003 |
| jawa | 0.6 | 0.004 | 0.005 | 0.006 | 0.007 | 0.003 |
| tengah | 0.6 | 0.004 | 0.005 | 0.008 | 0.003 | 0.003 |
| universitas | 0.6 | 0.004 | 0.005 | 0.017 | 0.003 | 0.003 |
| program | 0.6 | 0.004 | 0.005 | 0.014 | 0.003 | 0.003 |
| studi | 0.6 | 0.004 | 0.005 | 0.017 | 0.003 | 0.003 |
| lulus | 0.6 | 0.004 | 0.005 | 0.014 | 0.003 | 0.003 |
| didik | 0.6 | 0.004 | 0.005 | 0.062 | 0.003 | 0.006 |
| ahli | 0.6 | 0.004 | 0.005 | 0.014 | 0.003 | 0.003 |
| bidang | 0.6 | 0.004 | 0.005 | 0.02 | 0.003 | 0.003 |
| kerja | 0.6 | 0.004 | 0.005 | 0.028 | 0.003 | 0.003 |
| butuh | 0.6 | 0.004 | 0.005 | 0.008 | 0.003 | 0.009 |
| rektor | 0.6 | 0.004 | 0.005 | 0.014 | 0.003 | 0.003 |
| resmi | 0.6 | 0.004 | 0.005 | 0.011 | 0.01 | 0.003 |
| luncur | 0.6 | 0.004 | 0.005 | 0.006 | 0.017 | 0.006 |
| guru | 0.6 | 0.004 | 0.005 | 0.04 | 0.003 | 0.003 |
| sertifikasi | 0.6 | 0.004 | 0.005 | 0.023 | 0.003 | 0.003 |
| tenaga | 0.6 | 0.004 | 0.005 | 0.008 | 0.007 | 0.009 |
| pasar | 0.6 | 0.004 | 0.005 | 0.006 | 0.003 | 0.009 |
| negeri | 0.6 | 0.004 | 0.005 | 0.008 | 0.003 | 0.003 |
| swasta | 0.6 | 0.004 | 0.005 | 0.008 | 0.003 | 0.003 |
| sma | 0.6 | 0.004 | 0.005 | 0.014 | 0.003 | 0.003 |
| international | 0.6 | 0.004 | 0.005 | 0.011 | 0.01 | 0.003 |
| sekolah | 0.6 | 0.004 | 0.005 | 0.025 | 0.003 | 0.003 |

**Tabel 3.11** Hasil Hitung *Naïve Bayes* (Lanjutan)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Term** | **PVj** | P(Wk | Vj) | | | | |
| Politik | Olahraga | Pendidikan | Otomotif | Umum |
| pilih | 0.6 | 0.004 | 0.005 | 0.008 | 0.007 | 0.003 |
| smp | 0.6 | 0.004 | 0.005 | 0.011 | 0.003 | 0.003 |
| ajar | 0.6 | 0.004 | 0.005 | 0.017 | 0.003 | 0.014 |
| lahir | 0.6 | 0.004 | 0.005 | 0.006 | 0.003 | 0.012 |
| konsep | 0.6 | 0.004 | 0.005 | 0.006 | 0.007 | 0.003 |
| kait | 0.6 | 0.004 | 0.005 | 0.008 | 0.007 | 0.003 |
| rumah | 0.6 | 0.004 | 0.005 | 0.008 | 0.003 | 0.032 |
| kembang | 0.6 | 0.004 | 0.005 | 0.006 | 0.01 | 0.003 |
| damping | 0.6 | 0.004 | 0.005 | 0.006 | 0.003 | 0.012 |
| sistem | 0.6 | 0.004 | 0.005 | 0.008 | 0.007 | 0.003 |
| full | 0.6 | 0.004 | 0.005 | 0.011 | 0.007 | 0.003 |
| anak | 0.6 | 0.004 | 0.005 | 0.023 | 0.003 | 0.058 |
| orangtua | 0.6 | 0.004 | 0.005 | 0.017 | 0.003 | 0.003 |
| khawatir | 0.6 | 0.004 | 0.005 | 0.006 | 0.003 | 0.009 |
| bijak | 0.6 | 0.004 | 0.005 | 0.008 | 0.003 | 0.006 |
| proses | 0.6 | 0.004 | 0.005 | 0.006 | 0.003 | 0.009 |
| terampil | 0.6 | 0.004 | 0.005 | 0.008 | 0.003 | 0.006 |
| langsung | 0.6 | 0.004 | 0.005 | 0.006 | 0.01 | 0.003 |
| cari | 0.6 | 0.004 | 0.005 | 0.006 | 0.01 | 0.003 |
| pt | 0.6 | 0.004 | 0.005 | 0.003 | 0.014 | 0.003 |
| motor | 0.6 | 0.004 | 0.005 | 0.003 | 0.042 | 0.003 |
| giias | 0.6 | 0.004 | 0.005 | 0.003 | 0.017 | 0.003 |
| model | 0.6 | 0.004 | 0.005 | 0.003 | 0.028 | 0.006 |
| lampu | 0.6 | 0.004 | 0.005 | 0.003 | 0.035 | 0.003 |
| andal | 0.6 | 0.004 | 0.005 | 0.003 | 0.021 | 0.006 |
| imbang | 0.6 | 0.004 | 0.005 | 0.003 | 0.014 | 0.003 |
| roda | 0.6 | 0.004 | 0.005 | 0.003 | 0.014 | 0.003 |
| desain | 0.6 | 0.004 | 0.005 | 0.003 | 0.014 | 0.003 |
| hormat | 0.6 | 0.004 | 0.005 | 0.003 | 0.007 | 0.006 |
| harga | 0.6 | 0.004 | 0.005 | 0.003 | 0.028 | 0.003 |
| unit | 0.6 | 0.004 | 0.005 | 0.003 | 0.007 | 0.006 |
| booth | 0.6 | 0.004 | 0.005 | 0.003 | 0.014 | 0.003 |
| mobil | 0.6 | 0.004 | 0.005 | 0.003 | 0.021 | 0.003 |
| ktm | 0.6 | 0.004 | 0.005 | 0.003 | 0.024 | 0.003 |
| all | 0.6 | 0.004 | 0.005 | 0.003 | 0.021 | 0.003 |
| new | 0.6 | 0.004 | 0.005 | 0.003 | 0.021 | 0.003 |
| honda | 0.6 | 0.004 | 0.005 | 0.003 | 0.042 | 0.003 |
| beat | 0.6 | 0.004 | 0.005 | 0.003 | 0.042 | 0.003 |
| pakai | 0.6 | 0.004 | 0.005 | 0.003 | 0.042 | 0.003 |

**Tabel 3.11** Hasil Hitung *Naïve Bayes* (Lanjutan)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Term** | **PVj** | P(Wk | Vj) | | | | |
| Politik | Olahraga | Pendidikan | Otomotif | Umum |
| bohlam | 0.6 | 0.004 | 0.005 | 0.003 | 0.017 | 0.003 |
| hati | 0.6 | 0.004 | 0.005 | 0.003 | 0.007 | 0.012 |
| otomotif | 0.6 | 0.004 | 0.005 | 0.003 | 0.017 | 0.003 |
| lokasi | 0.6 | 0.004 | 0.005 | 0.003 | 0.007 | 0.006 |
| jangkau | 0.6 | 0.004 | 0.005 | 0.003 | 0.014 | 0.003 |
| fitur | 0.6 | 0.004 | 0.005 | 0.003 | 0.024 | 0.003 |
| timur | 0.6 | 0.004 | 0.005 | 0.003 | 0.007 | 0.006 |
| direktur | 0.6 | 0.004 | 0.005 | 0.003 | 0.01 | 0.003 |
| jaya | 0.6 | 0.004 | 0.005 | 0.003 | 0.007 | 0.006 |
| selatan | 0.6 | 0.004 | 0.005 | 0.003 | 0.007 | 0.026 |
| headlamp | 0.6 | 0.004 | 0.005 | 0.003 | 0.014 | 0.003 |
| led | 0.6 | 0.004 | 0.005 | 0.003 | 0.038 | 0.003 |
| tambah | 0.6 | 0.004 | 0.005 | 0.003 | 0.007 | 0.006 |
| rawat | 0.6 | 0.004 | 0.005 | 0.003 | 0.007 | 0.009 |
| arah | 0.6 | 0.004 | 0.005 | 0.003 | 0.007 | 0.006 |
| dukun | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.038 |
| bantu | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.014 |
| salin | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.017 |
| sehat | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.014 |
| dr | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.014 |
| peran | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.017 |
| genang | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.032 |
| banjir | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.023 |
| wib | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.046 |
| rt | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.014 |
| lurah | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.026 |
| cm | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.026 |
| camat | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.02 |
| barat | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.014 |
| rendam | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.017 |
| sentimeter | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.014 |
| rw | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.014 |
| pohon | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.014 |
| tumbang | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.014 |
| tangan | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.009 |
| tugas | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.006 |
| film | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.026 |
| kanker | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.017 |
| pers | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.006 |

**Tabel 3.11** Hasil Hitung *Naïve Bayes* (Lanjutan)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Term** | **PVj** | P(Wk | Vj) | | | | |
| Politik | Olahraga | Pendidikan | Otomotif | Umum |
| ketua | 0.6 | 0.013 | 0.01 | 0.003 | 0.003 | 0.003 |
| partai | 0.6 | 0.057 | 0.01 | 0.003 | 0.003 | 0.003 |
| sosok | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.009 |
| kabar | 0.6 | 0.009 | 0.005 | 0.003 | 0.003 | 0.003 |
| korupsi | 0.6 | 0.013 | 0.005 | 0.003 | 0.003 | 0.003 |
| asal | 0.6 | 0.004 | 0.005 | 0.003 | 0.01 | 0.003 |
| jelas | 0.6 | 0.004 | 0.005 | 0.003 | 0.007 | 0.003 |
| demokrat | 0.6 | 0.009 | 0.005 | 0.003 | 0.003 | 0.003 |
| dpp | 0.6 | 0.013 | 0.005 | 0.003 | 0.003 | 0.003 |
| sahabat | 0.6 | 0.004 | 0.01 | 0.003 | 0.003 | 0.003 |
| dewan | 0.6 | 0.009 | 0.005 | 0.003 | 0.003 | 0.003 |
| muhammad | 0.6 | 0.009 | 0.005 | 0.003 | 0.003 | 0.003 |
| tunggu | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.017 |
| golkar | 0.6 | 0.009 | 0.005 | 0.003 | 0.003 | 0.003 |
| tindak | 0.6 | 0.013 | 0.005 | 0.003 | 0.003 | 0.003 |
| sebut | 0.6 | 0.009 | 0.005 | 0.003 | 0.003 | 0.003 |
| beri | 0.6 | 0.022 | 0.005 | 0.006 | 0.007 | 0.003 |
| tugas | 0.6 | 0.004 | 0.01 | 0.014 | 0.003 | 0.003 |
| pidana | 0.6 | 0.017 | 0.005 | 0.003 | 0.003 | 0.003 |
| wakil | 0.6 | 0.013 | 0.01 | 0.014 | 0.003 | 0.003 |
| bambang | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.003 |
| tangkap | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.003 |
| komisi | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.003 |
| iii | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.003 |
| kpk | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.003 |
| dpr | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.003 |
| anggota | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.003 |
| fraksi | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.003 |
| giat | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.003 |
| konfirmasi | 0.6 | 0.004 | 0.005 | 0.003 | 0.003 | 0.003 |

Setelah mendapatkan nilai P(vj) dan P(wk | vj), hitung VMAP.

P(“tim”|vj) P(“kait”|vj) P(“tangan”|vj) P(“tugas”|vj) P(“pers”|vj) P(“ketua”|vj) P(“partai”|vj) P(“sosok”|vj) P(“kabar”|vj) P(“korupsi”|vj) P(“jelas”|vj) P(“demokrat”|vj) P(“dpp”|vj) P(“sahabat”|vj) P(“dewan”|vj) P(“muhammad”|vj) P(“tunggu”|vj) P(“golkar”|vj) P(“tindak”|vj) P(“sebut”|vj) P(“beri”|vj) P(“pidana”|vj) P(“wakil”|vj) P(“bambang”|vj) P(“tangkap”|vj) P(“komisi”|vj) P(“iii”|vj) P(“kpk”|vj) P(“dpr”|vj) P(“anggota”|vj) P(“fraksi”|vj) P(“fraksi”|vj) P(“giat”|vj) P(“konfirmasi”|vj)

* v”Politik” = 0.6\*0.004\*0.004\*0.004\*0.004\*0.004\*0.013\*0.057\*0.004\*0.009\*0.013 \*0.004\*0.009\*0.013\*0.004\*0.009\*0.009\*0.004\*0.009\*0.013\*0.009\*0.022\* 0.017\*0.013\*0.004\*0.004\*0.004\*0.004\*0.004\*0.004\*0.004\*0.004\*0.004\* 0.004

v”Politik” =3E-72

* v”Olahraga” = 0.6\*0.014\*0.005\*0.005\*0.005\*0.005\*0.01\*0.01\*0.005\*0.005\* 0.005\*0.005\*0.005\*0.005\*0.01\*0.005\*0.005\*0.005\*0.005\*0.005\*0.005\*0.005\*0.005\*0.01\*0.005\*0.005\*0.005\*0.005\*0.005\*0.005\*0.005\*0.005\*0.005\* 0.005

v”Olahraga” =7E-76

* v”Pendidikan” =0.6\*0.003\*0.008\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\* 0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\* 0.006\*0.003\*0.014\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\* 0.003\*0.003

v”Pendidikan” = 2E-83

* v”Otomotif” = 0.6\*0.003\*0.007\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\* 0.003\*0.007\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\* 0.007\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\* 0.003\*0.003

v”Otomotif” = 4E-81

* v”Umum” = 0.6\*0.003\*0.003\*0.009\*0.006\*0.006\*0.003\*0.003\*0.009\*0.003\* 0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.017\*0.003\*0.003\*0.003\* 0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\*0.003\* 0.003\*0.003

v”Umum” = 2E-82

Dari hasil perhitungan mencari nilai VMAPdidapat nilai tertinggi yaitu 3E-72. Jadi dokumen D1 termasuk ke dalam kategori **Politik**.

## Perancangan Sistem

Pada bagian ini akan dijelaskan mengenai perancangan sistem yang akan dikembangkan. Perancangan terdiri dari perancangan *use case diagram, class diagram, activity diagram, sequence diagram,* dan *statechart diagram.*

### *Use Case Diagram*

|  |
| --- |
|  |

**Gambar 3.6** *Use Case Diagram*

1. *Use Case* Melakukan *Scraping*

**Tabel 3.12** *Use Case* Melakukan *Scraping*

|  |  |
| --- | --- |
| *Usecase id* | UC-1 |
| *Usecase name* | Melakukan *Scraping* |
| *Actors* | *Website* |
| *Description* | *Usecase* ini menggambarkan proses pengambilan data dari halaman website seperti *link, image,* dan isi artikel. |
| *Precondition* | Aplikasi telah berjalan pada *web server* dan tersambung dengan jaringan internet. |
| *Normal flow* | 1. Sistem melakukan *load* dengan *link* beritayang sudah ditentukan. 2. Sistem akan melakukan proses *scraping* *website.* 3. Sistem menyimpan data hasil *scraping* ke dalam *database* sistem. |
| *Alternatif flow* | 1. Sistem melakukan *load* dengan *link* berita. 2. Sistem mengecek *link* berita. 3. Validasi gagal dan sistem tidak dapat melakukan *scraping website.* |

1. *Use Case* Melakukan *Preprocessing*

**Tabel 3.13** *Use Case*Melakukan *Preprocessing*

|  |  |
| --- | --- |
| *Usecase id* | UC-2 |
| *Usecase name* | Melakukan *Preprocessing* |
| *Actors* | *-* |
| *Description* | *Usecase* ini menggambarkan proses *preprocessing* isi website berupa teks yang terdiri dari *case folding, tokenizing, filtering,* dan *stemming.* |
| *Precondition* | Isi *web* hasil *scraping* yang berupa teks sudah tersimpan dalam *database* sistem. |
| *Normal flow* | 1. Sistem memanggil data yang tersimpan dalam *database* sistem. 2. Sistem melakukan proses *preprocessing.* |
| *Alternatif flow* | 1. Sistem memeriksa jumlah data dalam *database.* 2. Data dalam *database* tidak ada dan sistem tidak dapat melakukan proses *preprocessing.* |

1. Melakukan Proses Kategori

**Tabel 3.14** *Use Case* Melakukan Proses Kategori

|  |  |
| --- | --- |
| *Usecase id* | UC-3 |
| *Usecase name* | Melakukan Proses Kategori |
| *Actors* | *-* |
| *Description* | *Usecase* ini menggambarkan proses perhitungan teks pada dokumen data uji untuk mengetahui kategori yang sesuai dengan teks tersebut. |
| *Precondition* | Data teks sudah dilakukan proses *preprocessing.* |
| *Normal flow* | 1. Sistem melakukan perhitungan untuk mengetahui bobot setiap kata. 2. Sistem melakukan perhitungan untuk mengetahui kategori dokumen menggunakan algoritma *k-nearest neighbor.* 3. Sistem melakukan perhitungan untuk mengetahui kategori dokumen menggunakan algoritma *naïve bayes classifier.* |
| *Alternatif flow* | - |

1. Melakukan klasifikasi dengan knn

**Tabel 3.15** *Use Case* Melakukan klasifikasi dengan knn

|  |  |
| --- | --- |
| *Usecase id* | UC-4 |
| *Usecase name* | Melakukan klasifikasi dengan knn |
| *Actors* | *-* |
| *Description* | *Usecase* ini menggambarkan proses perhitungan teks pada dokumen data uji menggunakan algoritma *k-nearest neighbour* (knn) untuk mengetahui kategori yang sesuai dengan teks tersebut. |
| *Precondition* | Data teks sudah dilakukan proses *preprocessing.* |
| *Normal flow* | 1. Sistem melakukan perhitungan untuk mengetahui bobot setiap kata. 2. Sistem melakukan perhitungan untuk mengetahui kemiripan dokumen data uji dengan semua dokumen data latih yang sudah dikategorikan. |
| *Alternatif flow* | - |

1. Melakukan klasifikasi dengan *naïve bayes*

**Tabel 3.16** *Use Case* Melakukan klasifikasi dengan *naïve bayes*

|  |  |
| --- | --- |
| *Usecase id* | UC-4 |
| *Usecase name* | Melakukan klasifikasi dengan *naïve bayes* |
| *Actors* | *-* |
| *Description* | *Usecase* ini menggambarkan proses perhitungan teks pada dokumen data uji menggunakan algoritma *naïve bayes* untuk mengetahui kategori yang sesuai dengan teks tersebut. |
| *Precondition* | Data teks sudah dilakukan proses *preprocessing.* |
| *Normal flow* | 1. Sistem melakukan perhitungan untuk mencari nilai P(vj) 2. Sistem melakukan perhitungan untuk nilai P(wk | vj) 3. Sistem menghitung nilai probabilitas VMAP 4. Sistem mengambil nilai VMAP terbesar sebagai hasil dari klasifikasi. |
| *Alternatif flow* | - |

1. Melihat Kategori Teks Berita

**Tabel 3.17** *Use Case* Melihat Kategori Teks Berita

|  |  |
| --- | --- |
| *Usecase id* | UC-4 |
| *Usecase name* | Melihat Kategori Teks Berita |
| *Actors* | *Website Visitor* |
| *Description* | *Usecase* ini menggambarkan proses menampilkan halaman *user interface website* yang berisi artikel-artikel berita yang sudah dikategorikan dan halaman proses kategorisasi. |
| *Precondition* | Aplikasi sudah berjalan pada *web server* dan semua proses *usecase* berjalan sesuai fungsinya. |
| *Normal flow* | 1. Actor membuka halaman *website.* 2. Sistem menampilkan halaman *website* dengan artikel berita yang sudah dikategorikan. 3. Aktor membuka halaman *classification* untuk melihat proses pengkategorian dokumen. 4. Aktor menginputkan *link*  *website* untuk dilakukan perhitungan. 5. Sistem mengecek *link website* yang diinputkan. 6. Sistem melakukan proses *scraping, preprocessing,* menghitung bobot tiap kata, melakukan perhitungan untuk mendapatkan nilai kategori dengan algoritma knn dan *naïve bayes klassifier.* 7. Sistem menampilkan hasil dari proses-proses tersebut. |
| *Alternatif flow* | 1. Actor membuka halaman *website.* 2. Sistem menampilkan halaman *website* dengan artikel berita yang sudah dikategorikan. 3. Aktor membuka halaman *classification* untuk melihat proses pengkategorian dokumen. 4. Aktor menginputkan *link*  *website* untuk dilakukan perhitungan. 5. Sistem mengecek *link website* yang diinputkan. 6. Sistem gagal memvalidasi *link* yang diinputkan dan gagal melakukan proses pengkategorian. |

1. Melihat Grafik

**Tabel 3.18** *Use Case* Melihat Grafik

|  |  |
| --- | --- |
| *Usecase id* | UC-5 |
| *Usecase name* | Melihat Grafik |
| *Actors* | *Website Visitor* |
| *Description* | *Usecase* ini menggambarkan proses menampilkan halaman *user interface website* yang berisi grafik perbandingan proses kategori menggunakan *k-nearest neighbour* dan *naïve bayes.* |
| *Precondition* | Aplikasi sudah berjalan pada *web server* dan semua proses *usecase* berjalan sesuai fungsinya. |
| *Normal flow* | 1. Actor membuka halaman *website.* 2. Sistem menampilkan halaman *website* dengan artikel berita yang sudah dikategorikan. 3. Aktor membuka halaman Grafik untuk melihat grafik perbandingan proses kategori menggunakan *k-nearest neighbour* dan *naïve bayes.* 4. Sistem menampilkan grafik. |
| *Alternatif flow* | 1. Actor membuka halaman *website.* 2. Sistem menampilkan halaman *website* dengan artikel berita yang sudah dikategorikan. 3. Aktor membuka halaman Grafik untuk melihat grafik perbandingan proses kategori menggunakan *k-nearest neighbour* dan *naïve bayes.* 4. Sistem mengecek data proses kategori dalam *database.*. 5. Sistem gagal menampilkan grafik. |

### *Class Diagram*

### *Activity Diagram*

### *Sequence Diagram*

### *Statechart Diagram*

# DAFTAR PUSTAKA

|  |  |
| --- | --- |
| [1] | E. Prasetyo, Data Mining - Konsep dan Aplikasi Menggunakan MATLAB, Yogyakarta: C.V Andi Offset, 2012. |
| [2] | A. D. Arifin, I. Arieshanti and A. Z. Arifin, "Implementasi Algoritma K-Nearest Neighbour yang Berdasarkan One Pass Clustering untuk Kategorisasi Teks," p. 1. |
| [3] | Y. Samuel, R. Delima and A. Rachmat, "Implementasi Metode K-Nearest Neighbour dengan Decision Rule untuk Klasifikasi Subtopik Berita," *Jurnal Informatika,* pp. 1-15, 2014. |
| [4] | A. Hamzah, "Klasifikasi Teks dengan Naive Bayes Classifier (NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis," *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III,* pp. B-269 - 277, 2012. |
| [5] | S. L. B. Ginting and R. P. Trinanda, "Penggunaan Metode Naive Bayes Classifier pada Aplikasi Perpustakaan," pp. 1-7. |
| [6] | A. Hatta, N. Ramadijanti and A. Helen, "Rancang Bangun Sistem Pengelolaan Dokumen-Dokumen Penting Menggunakan Text Mining," pp. 1-4. |
| [7] | M. Salahudin and A. Rosa, Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek, Bandung: Informatika Bandung, 2013, p. 31. |
| [8] | Wikipedia, "Berita," 1 Maret 2016. [Online]. Available: https://id.wikipedia.org/wiki/Berita. |
| [9] | I. Sofana, Teori dan Modul Praktikum Jaringan Komputer, Bandung: Modula, 2011. |
| [10] | D. Juju, Kupas Tuntas CSS, Jakarta: PT. Elex Media Komputindo, 2007, p. 1. |
| [11] | Jayan, CSS untuk Orang Awam, Palembang: Maxikom, 2010. |
| [12] | A. C. Setiawan, Dahsyatkan Aplikasi PHP dengan Sentuhan Javascript, Yogyakarta: CV. Lokomedia, 2015. |
| [13] | B. Nugroho, Latihan Membuat Aplikasi Web PHP dan MySQL dengan Dreamweaver MX (6, 7, 2004) dan 8, Yogyakarta: Gava Media, 2009. |
| [14] | Jumadi and E. Winarko, "Penggunaan KNN (K-Nearest Neighbour) untuk Klasifikasi Teks Berita yang Tak-Terkelompokkan pada Saat Pengklasteran oleh STC (SUffix Tree Clustering)," pp. 50-81, 2015. |
| [15] | A. Sutanta, Sistem Basis Data, Yogyakarta: Graha Ilmu, 2004. |
| [16] | M. Shalahuddin and R. A.S, Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek, Bandung: Informatika Bandung, 2013. |

1. Shalahuddin, M., dan A.S, R, **“*Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek”***, Informatika Bandung*,* 2013, hal 29. [↑](#footnote-ref-1)
2. Shalahuddin, M., dan A.S, R, **“*Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek”***, Informatika Bandung*,* 2013, hal 29. [↑](#footnote-ref-2)