# 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 terjadinya pelimpahan informasi, sehingga menyebabkan manusia 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* yaitu 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.*

## *State of The Art*

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 1.1** *State of The Art*

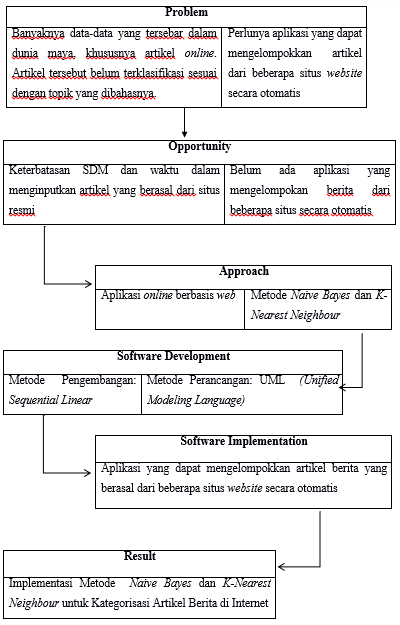
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Peneliti** | **Studi Kasus** | **Metode Pemecahan Masalah** | ***Tools* yang digunakan** | **Rancangan Sistem** |
| Arifin, dkk | *Dataset Reuters-21578* | Algoritma *K-Nearest Neighbour* berdasarkan *One Pass Clustering* | *Java Runtime Environtment 1.6.0.29* dan *XAMPP 1.7.7* | Perancangan sistem berbasis objek |
| Samuel, dkk | 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 | 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 |

**Tabel 1.1** *State of The Art* (Lanjut...)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Peneliti** | **Studi Kasus** | **Metode Pemecahan Masalah** | ***Tools* yang digunakan** | **Rancangan Sistem** |
| Hatta, dkk | umeDokn-dokumen penting yang berbentuk *hardcopy* | Algoritma *K-Nearest Neighbour* | *Java* | Perancangan sistem berbasis objek |
| Siti Nurpadilah | 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* |

## 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

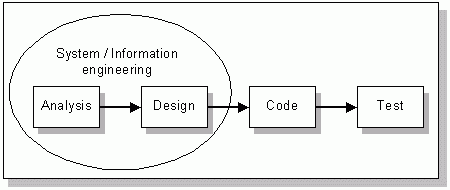
Studi pustaka di sini merupakan panduan dalam memecahkan masalah yang akan dibahas dengan mencari beberapa referensi sebagai landasan teoritis untuk menunjang penyusunan laporan tugas akhir ini, seperti buku cetak, jurnal, maupun referensi lainnya yang berasal dari internet.

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 teori-teori yang berhubungan dengan topik yang dibuat yang didasarkan oleh hasil penelitian dan hal-hal yang 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 uraian singkat teori-teori yang berhubungan mengenai topik yang akan dibuat. Teori-teori tersebut berdasarkan pada penelitian atau hal-hal lainnya yang didapatkan dari buku, jurnal maupun dari sumber lainnya. Teori-teori yang dijelaskan pada bab ini diantaranya yaitu artikel berita, internet, *website, scraping, 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.*

### 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]*.*

## 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. Terdapat beberapa tahapan yang dilakukan dalam *text mining* yaitu sebagai berikut:

1. *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).
2. *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.
3. *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.
4. *Stemming,* adalah proses mencari akar *(root)* kata dari setiap kata penyusun dokumen yaitu dengan mengembalikan kata berimbuhan ke dalam bentuk kata dasar.
5. *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 jarak setiap data latih dengan data uji. Untuk menghitung jarak data latih dengan data uji menggunakan formula jarak *Eucledian,* yaitu:

(3)

Keterangan:

*x* dan *y* adalah titik pada ruang vektor *n* dimensi.

*xi* dan *yi* adalah besaran scalar untuk dimensi ke-*I* dalam ruang vektor *n* dimensi.

* + - 1. Urutkan data latih berdasarkan jarak terdekat.
      2. Ambil data latih sebanyak nilai *k* yang sudah ditentukan.
      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. Bobot untuk setiap tetangga terdekat dihitung dengan rumus di bawah ini:

(4)

Formula yang dapat digunakan yaitu:

(5)

*v* merupakan jumlah data yang masuk ke dalam kelas *yi* [1].

## *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].

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.

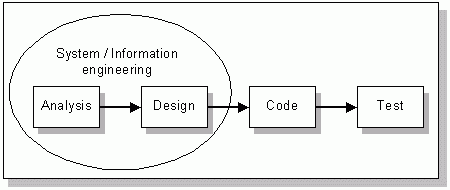
*nk* merupakan frekuensi kemunculan kata *wk* pada kategori *vj.*

*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.

### 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 extend* mendefinisikan hubungan antara dua *use case* dimana *use case* yang satu akan memanggil *use case* yang lain. |
| 5 | *Dependency extend* |  | *Dependency include* mendefinisikan pemanggilan *use case* lain yang hanya diperlukan pada kondisi tertentu. |
| 6 | Generalisasi |  | Generalisasi mendefinisikan hubungan antara dua *use case* dimana fungsi yang satu merupakan fungsi yang lebih umu dari fungsi yang lainnya. |

### 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, 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. |

### 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. |
| 3 | Percabangan / *decision* |  | Asosiasi percabangan digambarkan jika terdapat pilihan aktivitas yang lebih dari satu. |
| 4 | Penggabungan / *join* |  | Asosiasi penggabungan digambarkan jika terdapat beberapa aktivitas kemudian digabungkan menjadi satu. |
| 5 | Status akhir |  | Status akhir aktivitas yang dilakukan sistem. |
| 6 | *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. |

### 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. |

### 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 dank ode 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.

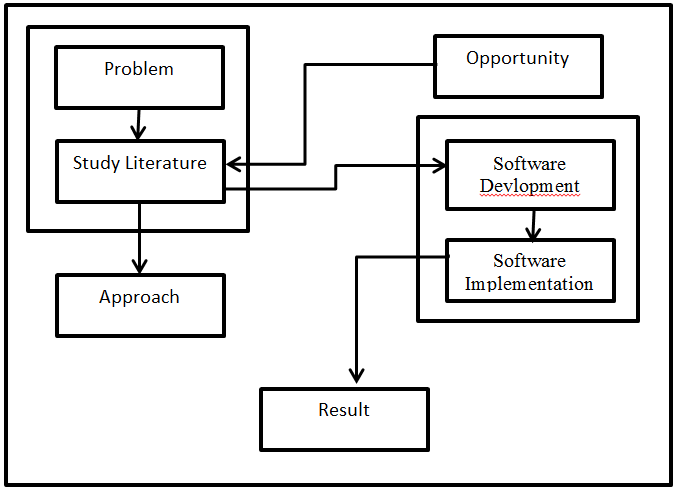
# BAB III

# 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, yaitu:

1. Tahap analisis kebutuhan perangkat lunak adalah tahap untuk mengumpulkan kebutuhan perangkat lunak yang dilakukan dengan intensif.
2. Tahap desain adalah tahap dimana kebutuhan perangkat lunak yang sebelumnya didapatkan pada tahap analisis akan digambarkan ke dalam desain perangkat lunak.
3. Tahap pengkodean 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*® *Core*™ i3-2310M CPU @ 2.10GHz 2.10GHz;
2. RAM 2 GB (1.89 GB usable);
3. *Harddisk* 464.3 GB;
4. *Monitor, keyboard, mouse* sebagai peralatan antarmuka.
5. Kebutuhan Perangkat Lunak *(Software)*

Pengembangan aplikasi dalam penelitian kali ini akan menggunakan perangkat lunak sebagai berikut:

1. Sistem operasi *Windows* 8 Pro;
2. *Browser internet*, seperti *Google Chrome* atau *Mozilla Firefox;*
3. *Software editor,* seperti *Sublime Text 2, notepad++, Adobe Dreamweaver CS3*;
4. Aplikasi *web server,* yaitu Apache versi 2.4.20;
5. Aplikasi *database server,* yaitu MySQL versi 5.6.21;
6. Perangkat lunak pemodelan sistem, seperti *StarUML* versi 5.0.2.1570
7. Perangkat lunak pemodelan *user interface,* seperti *Balsamiq Mockups*
8. Analisis Pengguna Aplikasi

Karakteristik pengguna yang akan berperan dalam penggunaan aplikasi ini yaitu hanya sebagai pemakai *(user)* saja.

### Analisis Kebutuhan Fungsional

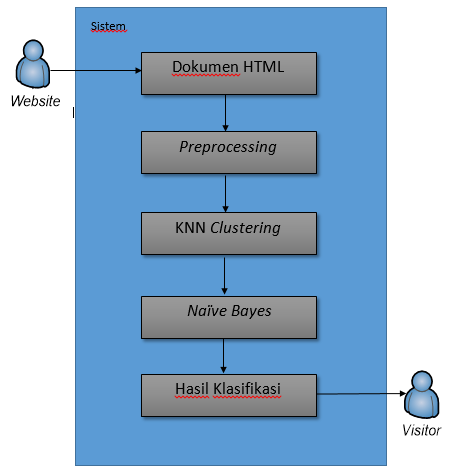
Analisis kebutuhan fungsional berisi mengenai deskripsi dan fitur-fitur yang terdapat pada aplikasi klasifikasi artikel berita *online* ini. Analisis kebutuhan fungsional meliputi deskripsi global dari aplikasi yang akan dikembangkan, arsitektural sistem yang akan dikembangkan, dan penggunaan metode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

# 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)