# BAB II

# LANDASAN TEORI

## Latar Belakang

Perkembangan teknologi di era globalisasi saat ini sangatlah pesat. Perkembangan teknologi ini menyebabkan banyaknya data-data yang tersebar di dunia maya. Penyebaran data-data tersebut bisa melalui melalui sms, media sosial, maupun melalui *website- website* yang ada di dunia internet.

Data-data yang tersebar dalam dunia maya bukan hanya sekitar 100 data atau 1000 data, melainkan jutaan atau mungkin bisa mencapai milyaran atau triliunan data. Data-data tersebut belum terklasifikasikan sesuai dengan topik apa yang dibahas dalam data tersebut. Hal ini akan membutuhkan waktu yang lama jika ingin melakukan pencarian sebuah data yang berisi mengenai satu topik tertentu. Maka dari itu diperlukanlah pengklasifikasian data sesuai dengan topik-topik yang dibicarakan oleh data-data tersebut.

Proses penglasifikasian dapat dilakukan dengan berbagai macam metode pengklasifikasian yang sebelumnya sudah dilakukan oleh para peneliti. Metode-metode itu 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.merdeka.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 |
| Hatta, dkk | umeDokn-dokumen penting yang berbentuk *hardcopy* | Algoritma *K-Nearest Neighbour* | *Java* | Perancangan sistem berbasis objek |

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Peneliti** | **Studi Kasus** | **Metode Pemecahan Masalah** | ***Tools* yang digunakan** | **Rancangan Sistem** |
| Siti Nurpadilah | Artikel berita dari detik.com, kompas.com, liputan6.com, viva.co.id, merdeka.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.

|  |  |
| --- | --- |
| **Problem** | |
| Banyaknya data-data yang tersebar dalam dunia maya, khususnya artikel *online*. Artikel tersebut belum terklasifikasi sesuai dengan topik yang dibahasnya. | Perlunya aplikasi yang dapat mengelompokkan artikel dari beberapa situs *website* secara otomatis |

|  |  |
| --- | --- |
| **Opportunity** | |
| Keterbatasan SDM dan waktu dalam menginputkan artikel yang berasal dari situs resmi | Belum ada aplikasi yang mengelompokan berita dari beberapa situs secara otomatis |

|  |  |
| --- | --- |
| **Approach** | |
| Aplikasi *online* berbasis *web* | Metode *Naive Bayes* dan *K-Nearest Neighbour* |

|  |  |
| --- | --- |
| **Software Development** | |
| Metode Pengembangan: *Sequential Linear* | Metode Perancangan: UML *(Unified Modeling Language)* |

|  |
| --- |
| **Software Implementation** |
| Aplikasi yang dapat mengelompokkan artikel berita yang berasal dari beberapa situs *website* secara otomatis |

|  |
| --- |
| **Result** |
| Implementasi Metode  *Naive Bayes* dan *K-Nearest Neighbour* untuk Kategorisasi Artikel Berita di Internet |

**Gambar 1.1** Kerangka Pemikiran

## Metode Penelitian

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

1. 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.

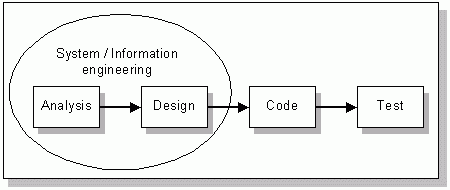
1. Studi pustaka

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.

1. 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.

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