

## BAB III

### METODOLOGI PENELITIAN

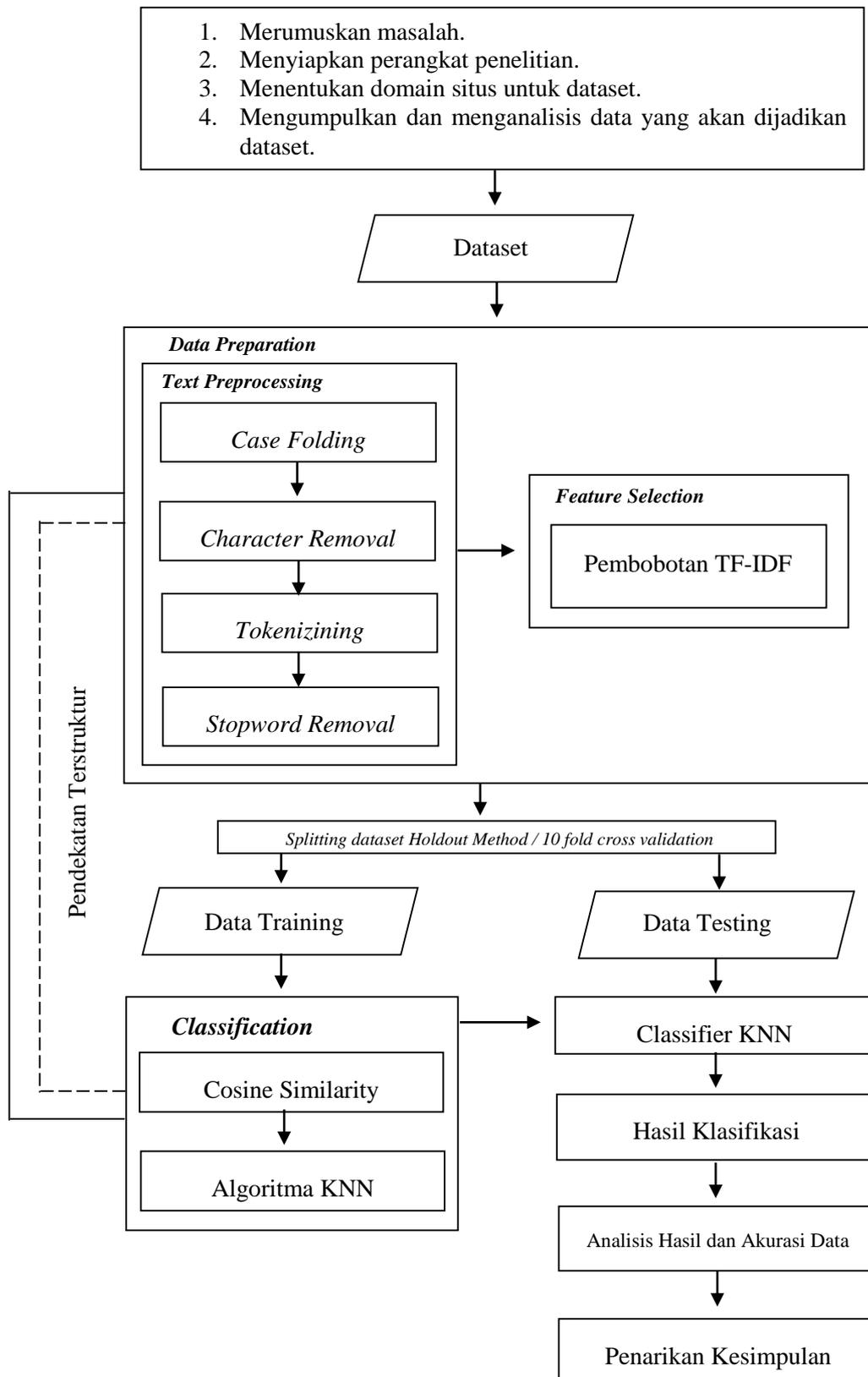
#### 3.1 Desain Penelitian

Langkah yang dilakukan setelah perumusan masalah yaitu membuat desain atau rancangan penelitian. Desain penelitian dapat memberikan strategi dalam memandu pengumpulan data, penganalisaan dan penginterpretasian data yang kemudian di dapatkan kesimpulan-kesimpulan. Penelitian ini meliputi beberapa tahapan penting diantaranya yaitu pengumpulan data set, perancangan *data preparation*, pengklasifikasian data dan analisis hasil pengujian. Pengembangan perangkat lunak dibuat secara terstruktur menggunakan metode *waterfall*. Berikut adalah *detail* tahapan penelitian yang digunakan.

1. Merumuskan masalah sebagai dasar penelitian dan menyiapkan perangkat penelitian untuk menunjang penelitian yang meliputi perangkat *hardware* maupun *software*.
2. Menentukan domain situs jurnal untuk mengambil kumpulan abstrak yang akan dikumpulkan sebagai dataset untuk bahan penelitian.
3. Dataset yang telah terkumpul di *input* ke dalam *file* excel csv sesuai kategorinya yang terdapat pada situs kemudian di *import* ke *database* dengan nama tabel *abstrak*.
4. Setelah data set di *import* ke *database* kemudian data akan melalui tahapan *data preparation* yang meliputi *preprocessing data* dan *feature selection* dengan pembobotan TF-IDF.
5. Tahapan awal pada *data preparation* yaitu melakukan *preprocessing data* atau pembersihan data yang meliputi *case folding*, *character removal*, *tokenizing* dan *stopword removal*. Hasil akhir dari *preprocessing data* di simpan pada *database* untuk di akses kembali pada tahapan pembobotan TF-IDF. Setelah dibersihkan, data hasil dari *preprocessing* melalui tahapan *feature selection* dengan pembobotan TF-IDF. Dan hasil dari pembobotan yaitu berupa *file* csv di mana *file* tersebut akan di akses kembali untuk digunakan pada tahapan klasifikasi data abstrak.

6. Setelah tahapan *data preparation* selesai maka data siap untuk digunakan pada tahapan klasifikasi. Sebelum di klasifikasikan data dibagi menjadi dua yaitu data *training* dan data *testing*. Dimana data *training* berfungsi dalam pembuatan bentuk classifier dengan algoritma KNN, sedangkan data *testing* berfungsi menguji classifier yang berbentuk memori dari data training yang telah dibuat..
7. Dalam pembagian data *training* dan data *testing*, penelitian ini menggunakan dua cara yaitu dengan *holdout method* dan metode *10 k-fold cross validation*. Pembagian data merupakan bagian dari skenario penelitian untuk menguji sistem klasifikasi yang dibuat. Pada metode *holdout* menggunakan 5 skenario yaitu : 80% data *training* – 20% data *testing*, 60% data *training* – 40% data *testing*, 50% data *training* – 50% data *testing*, 40% data *training* – 60% data *testing* dan 20% data *training* – 80% data *testing*. Sedangkan untuk metode pembagian data dengan *10 k-fold cross validation* terdapat 10 skenario dengan masing-masing 90% untuk data *training* dan 10% untuk data *testing*.
8. Setelah dipisahkan menjadi data *training* dan data *testing*, data melalui tahapan klasifikasi. Setiap skenario pada tahapan klasifikasi, data dilatih dengan menghitung jarak antar kedekatan dokumen *training* dengan *testing* yang menggunakan metode perhitungan jarak kedekatan yaitu *cosine similarity*. Hasil dari perhitungan jarak tersebut berbentuk vektor yang kemudian disimpan pada database.
9. Data yang telah melalui perhitungan jarak kemudian di panggil kembali untuk dilakukan tahapan klasifikasi menggunakan algoritma k-Nearest Neighbor dengan nilai tetangganya yaitu 3, 5, 7 dan 9.
10. Jika semua skenario telah di coba, maka data tersebut akan di analisis dengan menggunakan tabel *confusion matriks* untuk di cari nilai *precision*, *recall*, *f- measure* dan penarikan kesimpulan.

Dari detail tahapan yang dilakukan maka dibuatlah desain penelitian seperti yang terdapat pada Gambar 3.1.



### Gambar 3.1 Desain Penelitian

Desain penelitian yang dibuat pada Gambar 3.1 mengacu pada tahapan penemuan pengetahuan oleh Jiawei Han dalam *Data Mining Concepts and Techniques* tahun 2012 dan tahapan *text mining* oleh Tutkan dalam *Helmholtz Principle Based Supervised and Unsupervised Feature Selection Methods for Text Mining* tahun 2016. Tahapan dari keduanya telah di paparkan pada Bab 2, dan untuk alur prosesnya dapat dilihat pada Gambar 2.1 dan Gambar 2.2.

Secara garis besar bila melihat dari paparan alur pada Gambar 2.1 dan Gambar 2.2 untuk tahapan penemuan pengetahuan dan *text mining* memiliki tujuan yang sama yaitu pendistribusian dan pengolahan data, namun memiliki perbedaan dalam penamaan untuk setiap langkah alurnya. Dari segi proses penulis mengacu kepada kedua tahapan tersebut namun dari segi penamaan di dalam penelitian, penulis mengacu kepada tahapan *text mining*, karena penulis melakukan penelitian dalam pengklasifikasian data teks.

Tahapan *text mining* oleh Tutkan dalam *Helmholtz Principle Based Supervised and Unsupervised Feature Selection Methods for Text Mining* tahun 2016, terdiri dari pengumpulan dokumen, *text preprocessing*, *text transformation*, *features selection*, *data mining*, *interpretation* atau *evaluation*. Seperti pada Gambar 3.1 untuk langkah pertama penulis melakukan tahapan pengumpulan data. Kedua penulis melakukan persiapan data sebelum data tersebut diklasifikasikan dengan memberikannya istilah dengan *data preparation*. Langkah tersebut meliputi *text preprocessing* dan *feature selection*. Untuk penelitian yang dilakukan penulis tidak melakukan *text transformation* yaitu berupa *stemming* dikarenakan data atau *term* di dalam setiap dokumen sudah cukup baku dan terstruktur setelah dilakukan *stopword removal*. Kemudian pada proses *data mining* merupakan tahapan dalam pengklasifikasian data menggunakan algoritma klasifikasi. Langkah terakhir pada tahapan *text mining* adalah *interpretation* atau *evaluation* yang merupakan pengujian dan analisis terhadap hasil dari sistem klasifikasi yang telah dibuat.

### 3.1.1 Alat Penelitian

Merupakan perangkat yang digunakan untuk membuat aplikasi dan pengujian, terbagi menjadi 2 yaitu perangkat keras dan perangkat lunak, diantaranya,

1. Perangkat Keras
  1. Processor intel Dual Core N2840
  2. Random Access Memory (RAM) 2 GB DDR3
2. Perangkat Lunak
  1. XAMPP
  2. Microsoft Excel 2016
  3. Web Browser Google Chrome
  4. Netbeans IDE, Java JDK dan JRE
  5. Sublime Text 2
  6. Command Prompt

### 3.1.2 Bahan Penelitian

Bahan yang digunakan dalam penelitian ini yaitu data berupa.

1. Data abstrak pada jurnal internasional ilmu komputer dari situs Science Direct yang berjumlah 450 dengan 175-250 *term* pada setiap abstrak.
2. Membagi jurnal yang telah dikumpulkan menjadi 3 kelas kategori yaitu, *Computer in Education*, *Computer in Security* dan *Computer in Human Behavior*. Setiap kategori kelas memiliki jumlah abstrak sebanyak 150.

## 3.2 Metode Penelitian

Metode penelitian merupakan kegiatan dan prosedur yang digunakan untuk mendapatkan data dengan kegunaan tertentu untuk menghantarkan pada tujuan penelitian. Metode pada penelitian ini terbagi menjadi dua bagian yaitu metode pengumpulan data dan proses rekayasa sistem.

### 3.2.1 Metode Pengumpulan Data

Dalam tahapan pengumpulan data informasi yang sesuai dan dapat menunjang proses penelitian penulis menggunakan beberapa metode sebagai berikut :

#### 1. Metode Kepustakaan

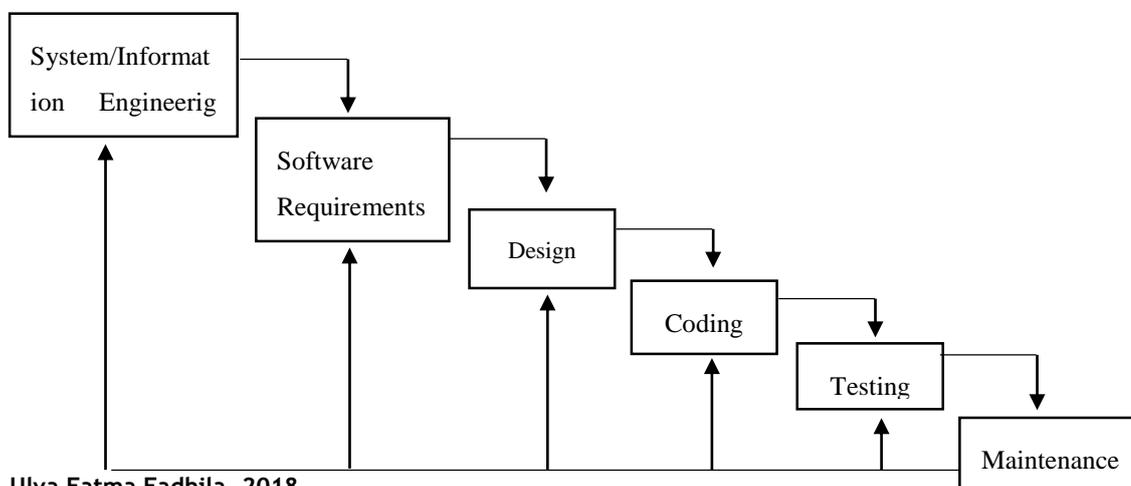
Melakukan pencarian sumber-sumber informasi dari buku, jurnal internasional dan nasional, *review paper*, berbagai literatur internet dan artikel yang berhubungan dengan penelitian. Hal tersebut dilakukan untuk mempelajari teori-teori yang menunjang di dalam penelitian.

#### 2. Mengumpulkan data set dari jurnal Ilmu Komputer

Mengumpulkan semua jurnal internasional dari situs Sciene Direct yang termasuk ke dalam 3 kategori yang telah ditentukan yaitu *Computer in Education*, *Computer in Security* dan *Computer in Human Behavior*. Pada proses pengumpulan data dilakukan dengan cara manual yaitu menyalin abstrak beserta kategorinya dari setiap jurnal ke dalam *file excel csv* yang kemudian di *input* ke dalam *database* untuk diproses.

### 3.2.2 Proses Pengembangan Perangkat Lunak

Pengembangan perangkat lunak dalam penelitian ini menggunakan model *waterfall*. Dengan model ini, akan dilakukan pendekatan secara sistematis dan terurut mulai dari analisis perangkat lunak, perancangan, implementasi, pengujian dan pemeliharaan. Secara umum, tahapan pada metode Waterfall dapat dilihat pada Gambar 3.2.



Ulva Fatma Fadhila, 2018

IMPLEMENTASI K-NEAREST NEIGHBOR DENGAN COSINE SIMILARITY UNTUK KLASIFIKASI ABSTRAK  
JURNAL INTERNASIONAL ILMU KOMPUTER

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Gambar 3.2 Metode *Waterfall* (Sukamto dan Shalahuddin, 2014)

Penjelasan dari Gambar 3.2 adalah sebagai berikut.

1. **Rekayasa dan Pemodelan Sistem (*System Engineering*)**  
Sistem merupakan bagian paling besar dalam pembangunan perangkat lunak, dimulai dengan membangun syarat dari semua elemen sistem dan mengalokasikan beberapa subsest dari kebutuhan ke software tersebut. Hal tersebut penting karena software harus berinteraksi dengan elemen lainnya seperti hardware, database dan sebagainya.
2. **Analisis Kebutuhan Perangkat Lunak (*Analysis*)**  
Pada tahap ini lebih difokuskan pada kebutuhan software. Hal tersebut meliputi pemahaman dari sifat program yang dibangun, tingkah laku, analisis dalam pemenuhan inforasi dan interface yang diperlukan. Dari kebutuhan sistem dan software harus didokumentasikan dan ditunjukkan kepada pelanggan.
3. **Perancangan Perangkat Lunak (*Design*)**  
Desain software merupakan langkah yang berfokus pada berbagai arsitektur program yang berbeda, yang meliputi struktur data, arsitektur software, representasi interface dan detail algoritma procedural.
4. **Pengkodean (*Coding*)**  
Pengkodean dilakukan agar sistem yang dibuat dimengerti oleh mesin, maka desain software harus diubah bentuknya menjadi bentuk bahasa pemrograman melalui proses coding.
5. **Pengujian Perangkat Lunak (*Testing*)**

Pengujian dilakukan untuk mengecek program (*software*) yang dibuat. Semua fungsi-fungsi software harus diujicobakan, agar software bebas dari error, dan hasilnya harus sesuai dengan kebutuhan yang telah didefinisikan.

6. Pemeliharaan Perangkat Lunak (*Maintenance*)

Perlunya pemeliharaan pada software karena mungkin saja memungkinkan masih ada error yang belum ditemukan. Pengembangan juga diperlukan ketika adanya perubahan dari eksternal perubahan seperti ketika ada pergantian sistem operasi, atau perangkat lainnya.