



## **PROPOSAL PENELITIAN**

### **Ekstraksi fitur statistik untuk deteksi noise pada dokumen Spesifikasi Kebutuhan Perangkat Lunak**

**Ahmad Mustofa  
NRP. 5116201054**

#### **DOSEN PEMBIMBING**

**Daniel Oranova Siahaan, S.Kom, M.Sc, P.D.Eng  
NIP. 197411232006041001**

#### **PROGRAM MAGISTER**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNOLOGI INFORMASI**

**INSTITUT TEKNOLOGI SEPULUH NOPEMBER**

**SURABAYA**

**2018**

*[Halaman ini sengaja dikosongkan]*

## **LEMBAR PENGESAHAN**

### **PROPOSAL TESIS**

Judul : Ekstraksi fitur statistik untuk deteksi noise pada dokumen  
Spesifikasi Kebutuhan Perangkat Lunak

Oleh : Ahmad Mustofa

NRP : 5116201054

Telah diseminarkan pada:

Hari :

Tanggal :

Tempat :

Mengetahui / menyetujui

Dosen Penguji:

1.

2.

3.

NIP.

Dosen Pembimbing:

1.

2.

*[Halaman ini sengaja dikosongkan]*

Nama Mahasiswa : Ahmad Mustofa

NRP : 5116201054

Pembimbing : Daniel Oranovora Siahaan, S.Kom. PD.Eng

## ABSTRAK

Tahap spesifikasi kebutuhan adalah tahap pertama yang dilakukan dalam proses pengembangan perangkat lunak. Sehingga jika terjadi kesalahan pada tahap ini, secara otomatis akan terjadi kesalahan pada tahap-tahap selanjutnya. Kesalahan dalam pernyataan kebutuhan perangkat lunak diantaranya berupa *noise* (derau), ambigu, konflik, serta inkonsistensi. Beberapa penelitian sebelumnya telah berhasil mendeteksi ambigu, konflik, dan inkonsistensi dalam pernyataan kebutuhan secara otomatis. Akan tetapi pada saat proposal ini ditulis, belum ada penelitian yang dilakukan untuk mendeteksi derau dalam pernyataan kebutuhan perangkat lunak secara otomatis.

Penelitian ini mengajukan suatu metode untuk deteksi derau dalam pernyataan kebutuhan perangkat lunak secara otomatis. Metode yang diajukan berfokus pada bagaimana melakukan ekstraksi fitur lokal dari masing-masing pernyataan kebutuhan dalam sebuah dokumen Spesifikasi Kebutuhan Perangkat Lunak (SKPL) sehingga kemudian bisa digunakan secara bersamaan dengan pernyataan kebutuhan dari dokumen SKPL yang lain dalam membangun model klasifikasi. Ekstraksi fitur lokal dari sebuah pernyataan kebutuhan dilakukan dengan memanfaatkan fitur statistik dari kemiripan pernyataan kebutuhan tersebut dengan pernyataan kebutuhan yang lain dalam dokumen SKPL yang sama.

Metode yang diajukan akan diuji dengan menggunakan data pernyataan kebutuhan yang telah dilabeli secara manual yang kemudian akan diukur performanya dengan menggunakan metode *k-fold cross validation*.

**Kata kunci:** pernyataan kebutuhan, deteksi derau, fitur statistik

*[Halaman ini sengaja dikosongkan]*

Nama Mahasiswa : Ahmad Mustofa

NRP : 5116201054

Pembimbing : Daniel Oranovora Siahaan, S.Kom. PD.Eng

## **ABSTRACT**

Requirement specification is the first step of a software development cycle. If errors occurred in this step, errors will automatically occur in the next step. Kesalahan dalam pernyataan kebutuhan perangkat lunak diantaranya berupa *noise* (derau), ambigu, konflik, serta inkonsistensi. Errors in software requirements consist of noise, ambiguous, conflict, and inconsistency. Some research has been successfully detecting ambiguous, conflict, and inconsistency in software requirements automatically. But when this document is written, there was no research done to detect noise in software requirements automatically.

This research proposes a method to detect noise in software requirements automatically. Proposed method focuses on how to extract local features of a requirement statement in a Software Requirement Specification (SRS) document so that this feature can be used globally with other requirement statements from another SRS document to build a classification model. Local feature extraction of a requirement statement is done by using statistical feature of the requirement statement's similarities with other requirement statement in the same SRS document.

Proposed method will be validated by using requirement statements data that have been labeled manually and its performance will be measured later by using k-fold cross validation method.

**Keywords:** requirements, noise detection, statistical feature

*[Halaman ini sengaja dikosongkan]*



## DAFTAR ISI

LEMBAR PENGESAHAN PROPOSAL TESIS.....	iii
ABSTRAK .....	v
ABSTRACT .....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xi
DAFTAR TABEL .....	xiii
1      BAB 1 PENDAHULUAN.....	1
1.1    Latar Belakang .....	1
1.2    Perumusan Masalah.....	2
1.3    Tujuan.....	2
1.4    Manfaat.....	2
1.5    Kontribusi Penelitian.....	2
1.6    Batasan Masalah.....	2
2      BAB 2 KAJIAN PUSTAKA .....	3
2.1    TF-IDF .....	3
2.2    Cosine Similarity .....	3
2.3    Support Vector Machine .....	3
2.4    Presisi dan Recall .....	6
3      BAB 3 METODOLOGI PENELITIAN .....	7
3.1    Pengumpulan dan Pelabelan Data .....	7
3.2    Studi Literatur.....	7
3.3    Praproses Data.....	7
3.3.1    Stopword Removal.....	7
3.3.2    Stemming .....	8
3.3.3    Tokenisasi.....	8
3.4    Penyiapan Fitur.....	8
3.4.1    Pembobotan Term .....	8
3.4.2    Ekstraksi Fitur .....	9
3.5    Penyusunan Model Klasifikasi.....	10
3.6    Pengujian dan Analisis .....	10
3.7    Penyusunan Laporan Penelitian .....	10

3.8	Jadwal Penelitian.....	10
4	DAFTAR PUSTAKA.....	11

## **DAFTAR GAMBAR**

GAMBAR 2.1 CONTOH ALTERNATIF HYPERPLANE .....	4
GAMBAR 3.1 ALUR PENELITIAN .....	7

*[Halaman ini sengaja dikosongkan]*

## DAFTAR TABEL

TABEL 3.1 CONTOH HASIL PENGHITUNGAN TERM FREQUENCY .....	9
TABEL 3.2 CONTOH HASIL PENGHITUNGAN INVERS DOCUMENT FREQUENCY .....	9
TABEL 3.3 CONTOH HASIL PENGHITUNGAN TF-IDF .....	9
TABEL 3.4 JADWAL KEGIATAN PENELITIAN TESIS .....	10

*[Halaman ini sengaja dikosongkan]*

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Tahap spesifikasi kebutuhan adalah tahap pertama yang dilakukan dalam proses pengembangan perangkat lunak. Sehingga jika terjadi kesalahan pada tahap ini, secara otomatis akan terjadi kesalahan pada tahap-tahap selanjutnya. Sudah banyak peneliti yang berusaha menemukan cara untuk mendeteksi kesalahan pada proses rekayasa kebutuhan. Bertrand meyer mengelompokkan kesalahan dalam spesifikasi kebutuhan menjadi tujuh kelompok yang kemudian dikenal dengan istilah *Meyer's seven sins* [1]. *Noise* (derau) adalah salah satu kesalahan dalam *Meyer's seven sins* yang disebabkan oleh adanya suatu elemen dalam teks yang memberikan informasi yang tidak relevan dengan domain masalah yang hendak diselesaikan [1]. Meyer membagi derau menjadi dua jenis, yaitu *remorse* dan *redundancy*. *Remorse* dapat didefinisikan sebagai pernyataan kebutuhan yang memberikan informasi yang tidak relevan dengan domain masalah, sedangkan *redundancy* adalah pernyataan kebutuhan yang memberikan informasi yang sama dengan pernyataan kebutuhan yang lain.

Di sisi lain, perkembangan yang pesat pada bidang *machine learning* telah mampu melakukan klasifikasi teks secara otomatis. Hal yang paling penting dalam pengolahan teks terletak pada tahap pembobotan. Hao Xu dan Bo Yu menggunakan metode pembobotan term frequency – invers document frequency (tf-idf) untuk melakukan klasifikasi email spam [2]. Bruno Trstenjak, Sasa Mikac, dan Dzenana Donko juga menggunakan tf-idf sebagai metode pembobotan untuk melakukan pengelompokan dokumen [3].

Derau dalam dokumen Spesifikasi Kebutuhan Perangkat Lunak (SKPL) hanya dapat dikenali secara lokal. Pernyataan kebutuhan yang tergolong derau dalam sebuah dokumen SKPL akan memiliki kriteria yang berbeda dengan pernyataan kebutuhan yang tergolong derau di dokumen yang SKPL lain. Hal ini bertolak belakang dengan metode tf-idf yang melakukan pembobotan teks secara global. Karenanya dibutuhkan sebuah metode untuk ekstraksi fitur lokal dari masing-masing pernyataan kebutuhan dalam sebuah dokumen SKPL.

Di sisi lain, B. Chandra, dan Manish Gupta menggunakan pendekatan statistik untuk seleksi fitur yang akan digunakan dalam klasifikasi gen [4]. Wujie Zhou, Lu Yu, Weiwei Qiu, Yang Zhou, dan Mingwei Wu juga menggunakan nilai statistik untuk mendapatkan fitur lokal dari sebuah citra yang kemudian digunakan untuk menentukan kualitas dari citra tersebut [5]. Pendekatan secara statistik inilah yang kemudian akan digunakan oleh penulis untuk ekstraksi fitur lokal dari masing-masing pernyataan kebutuhan dalam sebuah dokumen SKPL.

## **1.2 Perumusan Masalah**

Rumusan masalah yang diangkat dalam penelitian ini adalah:

1. Bagaimana menentukan fitur yang tepat
2. Bagaimana mengekstraksi fitur-fitur dari pernyataan kebutuhan dalam dokumen spesifikasi kebutuhan perangkat lunak

## **1.3 Tujuan**

Tujuan yang akan dicapai dalam pembuatan tesis ini adalah membangun sebuah metode yang dapat mendeteksi pernyataan kebutuhan yang termasuk derau dalam sebuah dokumen spesifikasi kebutuhan perangkat lunak.

## **1.4 Manfaat**

Manfaat dari penelitian ini adalah untuk membantu perekrut kebutuhan meningkatkan kualitas dokumen spesifikasi kebutuhan perangkat lunak yang akan dibuat dengan mengurangi pernyataan kebutuhan yang termasuk derau.

## **1.5 Kontribusi Penelitian**

Kontribusi yang diharapkan dari penelitian ini adalah mengembangkan sebuah metode ekstraksi fitur untuk klasifikasi pernyataan kebutuhan yang termasuk derau dalam dokumen spesifikasi kebutuhan perangkat lunak.

## **1.6 Batasan Masalah**

Batasan masalah pada penelitian ini adalah:

1. Dokumen spesifikasi kebutuhan perangkat lunak yang digunakan hanya menggunakan Bahasa Inggris.



## BAB 2

### KAJIAN PUSTAKA

#### 2.1 TF-IDF

Term frequency – invers document frequency adalah metode pembobotan dalam pengolahan teks. Nilai term frequency (tf) didapat dengan menghitung jumlah kemunculan sebuah term dalam sebuah dokumen. Sedangkan nilai invers document frequency didapat dengan rumus

$$\text{idf}_i = \log_2 \frac{N}{\text{df}_i} \quad (2.1)$$

dimana  $\text{df}_i$  adalah jumlah dokumen yang mengandung term  $i$ . Nilai tf-idf dari term  $i$  kemudian dapat dihitung dengan mengalikan nilai tf dari term  $i$  dengan nilai idf dari term  $i$  [3].

#### 2.2 Cosine Similarity

Dalam *text processing*, sebuah dokumen biasanya direpresentasikan sebagai sebuah vektor. Cosine similarity adalah sebuah metode yang digunakan untuk menghitung kemiripan antara dua vektor. Nilai cosine similarity dari vektor A dan vektor B didapat dengan menggunakan rumus

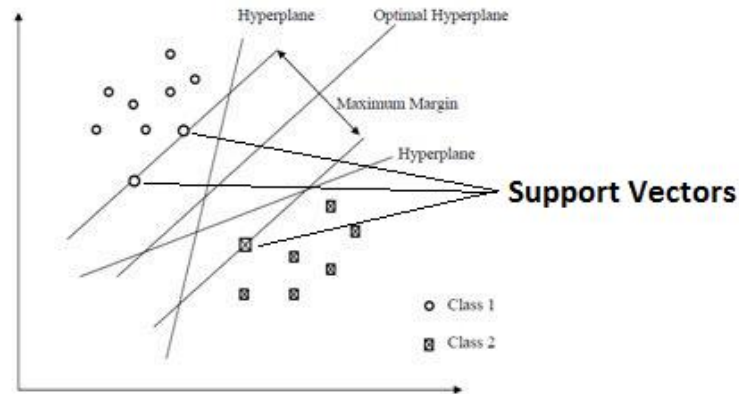
$$\text{sim}(A, B) = \frac{\sum_{k=1}^N A_k * B_k}{\sqrt{\sum_{k=1}^N A_k^2} * \sqrt{\sum_{k=1}^N B_k^2}} \quad (2.2)$$

dimana  $A_k$  dan  $B_k$  adalah komponen dari vektor A dan B [6].

#### 2.3 Support Vector Machine

*Support vector machine* (SVM) adalah metode klasifikasi *supervised* (data latih sudah diketahui kelasnya) yang mengklasifikasikan dua kelas. Pada metode ini, setiap data akan di-*plotting* ke dalam ruang  $n$ -dimensi (dimana  $n$  adalah jumlah fitur) dengan nilai masing-masing fitur menjadi nilai tertentu pada koordinat. Kemudian akan dilakukan klasifikasi dengan mencari *hyperplane* (bidang pembatas) untuk membedakan antara dua kelas sebaik mungkin. Gambar 2.1 merupakan beberapa contoh dari kemungkinan *hyperplane* yang digunakan untuk memisahkan kelas satu dengan lainnya. Metode SVM

akan mencari *hyperplane* yang memiliki *margin* yang maksimal. *Support vector* adalah data-data yang dijadikan pembatas dengan kelas lain.



Sumber gambar : [www.sine.ni.com](http://www.sine.ni.com)

Gambar 2.1 Contoh alternatif hyperplane

Diberikan data masukan  $(x_1, x_2, x_3, \dots, x_n)$  dan masing-masing kelas dinotasikan  $y_i \in \{-1, +1\}$  untuk  $i = 1, 2, 3, \dots, n$  dimana  $n$  adalah banyaknya data. Fungsi *hyperplane* dibuat dengan persamaan

$$w \cdot x + b = 0 \quad (2.3)$$

dengan batasan yang ditulis dalam persamaan

$$y_i(wx_i + b) + t_i \geq 1 \quad (2.4)$$

untuk mencari nilai  $w$  dan  $b$  yang optimal digunakan persamaan

$$\min \frac{1}{2} ||w||^2 + C \sum_{i=1}^l t_i \quad (2.5)$$

dimana nilai  $C$  adalah nilai pinalti dari kesalahan klasifikasi. Fungsi tujuan persamaan di atas berbentuk kuadrat, sehingga untuk menyelesaikannya, bentuk tersebut ditransformasikan ke dalam bentuk *dual space*. Persamaan *dual space* dapat ditulis menggunakan persamaan

$$\max \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j x_i x_j \quad (2.6)$$

dengan batasan dalam persamaan berikut.

$$\alpha_i \geq 0, \sum_{i=1}^l \alpha_i y_i = 0 \quad (2.7)$$

Untuk mencari nilai  $\alpha_i$ , digunakan *sequential minimal optimization* (SMO). Persamaan *hyperplane* dilakukan dengan persamaan:

$$f = w^T z + b = \sum_{i=1}^s \alpha_i y_i x_i^T z + b \quad (2.8)$$

dimana  $z$  adalah data masukan *support vector machine*. Pada banyak kasus, data yang diklasifikasikan tidak bisa langsung dipisahkan dengan garis yang linear. Oleh karena itu, digunakan metode kernel untuk mengatasi permasalahan tersebut. Dengan metode kernel, suatu data  $x$  di *input space* dipetakan ke fitur *space F* dengan dimensi yang lebih tinggi. Salah satu kernel yang biasa dipakai adalah kernel RBF dan *Polynomial*. Persamaan kernel RBF dan *Polynomial* berurutan dapat dilihat pada persamaan di bawah ini:

$$k(x, y) = \exp(-\gamma |x - y|^2) \quad (2.9)$$

$$k(x, y) = (\gamma \langle x^T y \rangle + r)^p \quad (2.10)$$

Penggunaan fungsi kernel mengubah persamaan *dual space* menjadi

$$\max \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j k(x_i, x_j) \quad (2.11)$$

dan juga mengubah persamaan *hyperplane* menjadi

$$f = \sum_{i=1}^s \alpha_i y_i k(x_i, z) + b \quad (2.12)$$

[7].

## 2.4 Presisi dan Recall

Untuk menghitung tingkat performansi suatu sistem dapat digunakan perhitungan presisi dan *recall*. Secara matematis, rumus untuk menghitung presisi dan recall dapat dilihat pada persamaan berikut

$$precision = \frac{TP}{TP + FP} \quad (2.13)$$

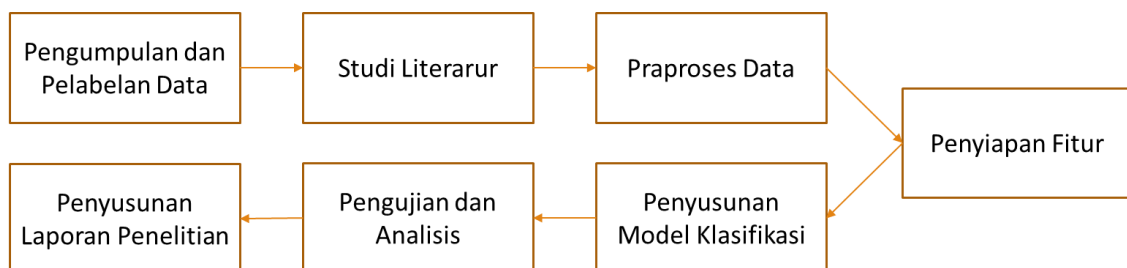
$$recall = \frac{TP}{TP + FN} \quad (2.14)$$

dimana TP adalah kebutuhan yang tergolong derau dan terklasifikasi sebagai derau, FP adalah kebutuhan yang bukan tergolong derau akan tetapi terklasifikasi sebagai derau, serta FN adalah kebutuhan yang tergolong derau akan tetapi tidak terklasifikasi sebagai derau.

## BAB 3

### METODOLOGI PENELITIAN

Bab ini akan memaparkan tentang metodologi penelitian yang digunakan pada penelitian ini, yang terdiri dari (1) pengumpulan dan pelabelan data, (2) studi literatur, (3) praproses data, (4) ekstraksi fitur, (5) penyusunan model klasifikasi, (6) pengujian dan analisis, dan (7) penyusunan laporan penelitian. Ilustrasi alur metodologi penelitian dapat dilihat pada berikut.



Gambar 3.1 Alur Penelitian

#### 3.1 Pengumpulan dan Pelabelan Data

Penelitian diawali dengan proses pengumpulan data pernyataan kebutuhan yang kemudian dilakukan pelabelan oleh beberapa pakar untuk menentukan pernyataan kebutuhan yang tergolong derau atau tidak.

#### 3.2 Studi Literatur

Proses kedua adalah pengkajian yang berkaitan dengan topik penelitian yang diambil. Pada penelitian ini, referensi yang digunakan adalah jurnal-jurnal yang berkaitan dengan metode-metode yang digunakan.

#### 3.3 Praproses Data

Pada tahap ini akan dilakukan praproses pada data pernyataan kebutuhan yang telah berhasil dikumpulkan. Tahap ini terbagi menjadi 3 proses, yaitu *stopword removal*, *stemming*, dan tokenisasi.

##### 3.3.1 Stopword Removal

*Stopword removal* adalah penghapusan kata penghubung dan tanda baca dari masing-masing data pernyataan kebutuhan. Sebagai contoh, sebuah pernyataan

kebutuhan “show information and availability of lecturer” akan menjadi “show information availability lecturer” setelah melalui tahap ini.

### **3.3.2 Stemming**

*Stemming* adalah proses pengubahan setiap kata dalam pernyataan kebutuhan menjadi kata dasar. Sebagai contoh, “show information availability lecturer” akan menjadi “show information available lecture” setelah melalui tahap ini.

### **3.3.3 Tokenisasi**

Tokenisasi adalah proses pemecahan masing-masing pernyataan kebutuhan menjadi kumpulan kata yang unik. Sebagai contoh, “show information available lecture available student” akan menjadi {“show”, “information”, “available”, “lecture”, “student”} setelah melalui tahap ini.

## **3.4 Penyiapan Fitur**

Tahap ini terdiri dari dua proses utama, yaitu tahap pembobotan term dan tahap ekstraksi fitur.

### **3.4.1 Pembobotan Term**

Masing-masing pernyataan kebutuhan yang telah melalui tahap praproses akan dihitung bobotnya dengan menggunakan rumus tf-idf. Setelah proses ini, masing-masing pernyataan kebutuhan akan direpresentasikan sebagai sebuah vektor.

Sebagai contoh, 2 pernyataan kebutuhan F01 “show information and availability of lecturer” dan F02 “update lecturer’s account information” akan menjadi {“show”, “information”, “available”, “lecture”} dan {“update”, “lecture”, “account”, “information”} setelah melalui tahap praproses. Token-token dari F01 dan F02 kemudian digabungkan menjadi {“show”, “information”, “available”, “lecture”, “update”, “account”} untuk kemudian digunakan sebagai fitur dalam penghitungan bobot tf-idf.

Untuk mendapatkan bobot tf-idf, pertama yang dilakukan adalah menghitung *term frequency* dari masing-masing kata fitur dalam masing-masing pernyataan kebutuhan sebagai berikut.

Tabel 3.1 Contoh Hasil Penghitungan Term Frequency

Kode	Show	information	available	lecture	update	Account
F01	1	1	1	1	0	0
F02	0	1	0	1	1	1

Setelah itu, dilakukan penghitungan idf dari masing-masing kata fitur sebagai berikut.

Tabel 3.2 Contoh Hasil Penghitungan Invers Document Frequency

Kode	Show	information	available	lecture	update	Account
Idf	0.3	0	0.3	0	0.3	0.3

Tahap akhir adalah mengalikan nilai tf dari masing-masing pernyataan kebutuhan dengan nilai idf.

Tabel 3.3 Contoh Hasil Penghitungan TF-IDF

Kode	show	information	available	lecture	update	Account
F01	0.3	0	0.3	0	0	0
F02	0	0	0	0	0.3	0.3

### 3.4.2 Ekstraksi Fitur

Untuk masing-masing vektor pernyataan kebutuhan, akan dihitung nilai similaritasnya dengan vektor pernyataan kebutuhan yang lain dalam dokumen SKPL yang sama dengan menggunakan *cosine similarity*. Setelah itu diambil tiga fitur yang kemudian akan digunakan dalam tahap klasifikasi. Tiga fitur tersebut adalah sebagai berikut.

a. Maximum of Similarity Measure

Fitur ini adalah nilai maksimum dari nilai similaritas yang sudah didapatkan sebelumnya.

b. Mean of Similarity Measure

Fitur ini adalah nilai rata-rata dari nilai similaritas yang sudah didapatkan sebelumnya.

c. Standard Deviation of Similarity Measure

Fitur ini adalah nilai standar deviasi dari nilai similaritas yang sudah didapatkan sebelumnya.

Berikut adalah pseudocode dari proses ekstraksi fitur yang dilakukan untuk masing-masing pernyataan kebutuhan dalam sebuah dokumen SKPL.

1	<b>ExtractFeature (DokumenSKPL)</b>
2	feature_matrix = new matrix
3	foreach k = pernyataan_kebutuhan in DokumenSKPL
4	similarity = cosineSimilarity(k, rest_of_pernyataan_kebutuhan)
5	feature_matrix.add([mean(similarity), std(similarity), max(similarity)])
6	end foreach
7	return feature_matrix

### 3.5 Penyusunan Model Klasifikasi

Pada tahap ini data-data pernyataan kebutuhan yang sudah direpresentasikan sebagai vektor fitur akan dibagi menjadi 2 kelompok, yaitu data latih dan data uji. Data latih akan digunakan untuk menyusun model klasifikasi yang kemudian akan digunakan untuk mengklasifikasi data uji.

### 3.6 Pengujian dan Analisis

Pada tahap ini akan dilakukan pengklasifikasian data uji dengan menggunakan model klasifikasi yang sudah terbentuk sebelumnya. Hasil klasifikasi dari data uji ini kemudian akan dianalisa performanya sehingga dapat diketahui apakah metode yang diusulkan dapat diterima atau tidak.

### 3.7 Penyusunan Laporan Penelitian

Pada tahap ini, semua hasil penelitian akan didokumentasikan dalam bentuk laporan penelitian.

### 3.8 Jadwal Penelitian

Jadwal penelitian yang dilakukan dapat dilihat pada tabel berikut

Tabel 3.4 Jadwal Kegiatan Penelitian Tesis

No.	Kegiatan	Bulan															
		Maret 2018				April 2018				Mei 2018				Juni 2018			
1.	Pengumpulan dan Pelabelan Data																
2.	Studi Literatur																
3.	Praproses Data																
4.	Penyiapan Fitur																
5.	Penyusunan Model Klasifikasi																
6.	Pengujian dan Analisis																
7.	Penyusunan Laporan Penelitian																



## DAFTAR PUSTAKA

- [1] B. Meyer, "On Formalism in Specification," *IEEE Software*, 1985.
- [2] H. Xu and B. Yu, "Automatic thesaurus construction for spam filtering using revised back propagation neural network," *Expert Systems with Applications*, no. 37, pp. 18-23, 2010.
- [3] B. Trstenjak, S. Mikac and D. Donko, "KNN with TF-IDF Based Framework for Text Categorization," *Procedia Engineering*, no. 69, pp. 1356-1364, 2014.
- [4] B. Chandra and M. Gupta, "An efficient statistical feature selection approach for classification of gene," *Journal of Biomedical Informatics*, vol. 44, pp. 529-535, 2011.
- [5] W. Zhou, L. Yu, W. Qiu, Y. Zhou and M. Wu, " Local Gradient Patterns (LGP): An Effective Local-Statistical-Feature Extraction Scheme for No-Reference Image Quality Assessment," *Information Sciences*, Vols. 397-398, pp. 1-14, 2017.
- [6] A. Singhal, "Modern Information Retrieval: A Brief Overview," Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2001.
- [7] J. C. Platt, "Sequential Minimal Optimization : A Fast Algorithm for Training Support Vector Machines," Technical Report MSR-TR-98-14, 1998.