**Analisis Sentimen Terhadap Tanggapan Mahasiswa Tentang Layanan Pusat Bahasa**

**(Studi Kasus : Telkom University)**

**Tugas Akhir**

**diajukan untuk memenuhi salah satu syarat**

**memperoleh gelar sarjana**

**dari Program Studi Teknik Informatika**

**Fakultas Informatika**

**Universitas Telkom**

**1301168399**

**Adi Satria Pangestu**

****

**Program Studi Sarjana Teknik Informatika**

**Fakultas Informatika**

**Universitas Telkom**

**Bandung**

**2018**

**LEMBAR PENGESAHAN**

**Analisis Sentimen Terhadap Tanggapan Mahasiswa Tentang Layanan Pusat Bahasa**

**(Studi Kasus : Telkom University)**

**Sentiment Analysis Against Student Feedback About the Language Center Service (Case Study : Telkom University)**

**NIM : 1301168399**

**Adi Satria Pangestu**

Tugas akhir ini telah diterima dan disahkan untuk memenuhi sebagian syarat memperoleh gelar pada Program Studi Sarjana Teknik Informatika  
Fakultas Informatika

Universitas Telkom

Bandung, 4/4/2018

Menyetujui

|  |  |  |
| --- | --- | --- |
| Pembimbing I, |  | Pembimbing II, |
|  |  |  |
| Yuliant Sibaroni, S.Si.,M.T  00750036 |  | Mohamad Syahrul Mubarok  10830028 |

|  |
| --- |
| Ketua Program Studi Sarjana Teknik Informatika,  Said Al Faraby, S.T, M.Sc.  NIP: 15980019 **LEMBAR PERNYATAAN** Dengan ini saya, Adi Satria Pangestu, menyatakan sesungguhnya bahwa Tugas Akhir saya dengan judul Analisis Sentimen Terhadap Tanggapan Mahasiswa Tentang Layanan Pusat Bahasa (Studi Kasus : Telkom University) beserta dengan seluruh isinya adalah merupakan hasil karya sendiri, dan saya tidak melakukan penjiplakan yang tidak sesuai dengan etika keilmuan yang belaku dalam masyarakat keilmuan. Saya siap menanggung resiko/sanksi yang diberikan jika di kemudian hari ditemukan pelanggaran terhadap etika keilmuan dalam buku TA atau jika ada klaim dari pihak lain terhadap keaslian karya,  Bandung, <Tanggal/bulan/tahun>  Yang Menyatakan  Adi Satria Pangestu |

**Analisis Sentimen Terhadap Tanggapan Mahasiswa Tentang Layanan Pusat Bahasa**

**(Studi Kasus : Telkom University)**

AdiSatriaPangestu1

1,2,3Fakultas Informatika, Universitas Telkom, Bandung

4Divisi Digital Service PT Telekomunikasi Indonesia

1adisatriaa@students.telkomuniversity.ac.id, 2 yuliant@telkomuniversity.ac.id, 3 msyahrulmubarok@telkomuniversity.ac.id

#### Abstrak

Pendidikan begitu penting bagi setiap orang dan dunia pendidikan tidak bisa lepas dari kehidupan sehari-hari. Banyak instansi yang bergerak dibidang pendidikan salah satunya Telkom University. Perkembangan instansi pendidikan dapat dipengaruhi oleh beberapa faktor salah satunya layanan pusat bahasa di Telkom University. Untuk meningkatkan pelayanan yang ada, diperlukannya tanggapan-tanggapan dan opini mahasiswa terkait layanan tersebut. Dari opini tersebut dapat dimanfaatkan untuk mengetahui kelebihan dan kekurangan suatu layanan. Tahap pengambilan kesimpulan opini ini menggunakan sentiment analys. Sistem yang dibangun terdiri dari 3 bagian besar yaitu *pre-processing* data, feature extraction, proses klasifikasi.

*Pre-processing* data dilakukan untuk membersihkan kata dari setiap kalimat menggunakan tahap cleansing, case folding, filtering, lemmatization, dan convert negation. Setelah dilakukan *pre-processing* maka akan dilakukan feature extraction menggunakan POS Tagging untuk pemberian label jenis kata pada setiap kata di dalam kalimat apakah termasuk kata benda, kata kerja, kata sifat, dan kata keterangan sesuai keluaran yang diharapkan. Metode klasifikasi yang digunakan dalam tugas akhir ini menggunakan Jaringan Saraf Tiruan dan Naïve Bayes, sehingga pada tugas akhir ini akan dilakukan perbandingan antara kedua klasifikasi tersebut untuk mencari nilai akurasi yang terbaik.

**Kata kunci : layanan pusat bahasa, sentiment analys, pre-processing, naïve bayes, jaringan saraf tiruan.**

# **Abstract**

**Education is so important to everyone and the world of education can not be separated from everyday life. Many agencies are engaged in education one of Telkom University. The development of educational institutions can be influenced by several factors, one of which is the language center service at Telkom University. To improve existing services, students' responses and opinions regarding the service are required. From these opinions can be utilized to know the advantages and disadvantages of a service. The conclusion of this opinion stage using the sentiment analys. The built system consists of 3 major parts of *pre-processing* data, feature extraction, classification process.**

***Pre-processing* data is done to clear the word from every sentence using cleansing stage, case folding, filtering, lemmatization, and convert negation. After *pre-processing* it will be done feature extraction using POS Tagging for labeling the type of word on every word in the sentence whether including nouns, verbs, adjectives, and adverbs according to the expected output. The classification method used in this final project is using Artificial Neural Network and Naïve Bayes, so in this final project we will compare the two classification to find the best accuracy value.**

**Keywords: sentiment analysis, artificial neural network, naïve bayes, Information System Services.**

1. **Pendahuluan**

Pendidikan telah menumbuhkan banyak instansi yang semakin berkembang di Indonesia, terutama instansi yang bergerak di bidang jasa pendidikan salah satunya Universitas Telkom. Setiap instansi akan berusaha untuk berlomba-lomba di dalam meningkatkan mutu pelayanannya, salah satu layanan di Universitas Telkom yaitu Layanan Pusat Bahasa.

Setiap instansi akan mendapatkan dampak dari hasil persaingan-persaingan yang muncul dan membuat instansi dapat meningkatkan pelayanan yang telah ada selama ini. Peningkatan layanan bertujuan sebagai wujud respon kepada mahasiswa/i di generasi yang akan datang, sehingga diharapkan pelayanan sebelumnya bisa menjadi lebih baik lagi. Maka dari itu, instansi harus menganggap setiap perubahan, persaingan dan ketidakpastian dari setiap permintaaan konsumen sebagai acuan untuk terus berkembang.

Berdasarkan hal tersebut, penelitian ini melakukan analisis sentimen terhadap tanggapan mahasiswa tentang Layanan Pusat Bahasa di Universitas Telkom demi meningkatkan mutu layanannya. Data yang dianalisi tersebut berupa text yang bernilai positif dan negatif dan dapat diolah dengan analisis sentimen. Penggunaan analisis sentimen dapat digunakan untuk mengetahui opini yang mengungkapkan persentase pandangan positif dan negatif.

Pada tugas akhir ini akan menerapkan metode klasifikasi jaringan saraf tiruan dan multinomial naïve bayes untuk melakukan perbandingan antara dua classifier tersebut. Penggunaan metode klasifikasi jaringan saraf tiruan ini diharapkan dapat memberikan tingkat keakuratan yang tinggi.

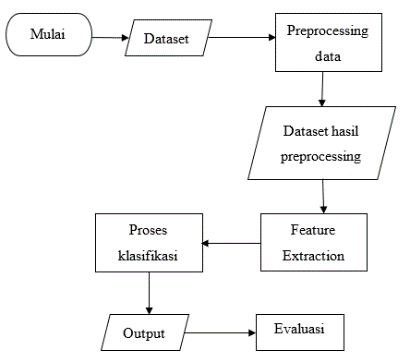
1. **Studi Terkait**

Dalam penelitian ini menggunakan beberapa referensi sebelumnya yang terkait dengan beberapa penelitian. Penelitian yang terkait diantaranya Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet. Fokus penelitian ini adalah untuk menemukan kelas positif dan negatif dari data retweet sejumlah 400. Pembobotan teks menggunakan TFIDF dan menggunakan pembobotan fitur jumlah *retweet* (non-tekstual). Data tersebut diambil dari hasil retweet opini masyarakat terhadap tayangan televisi di twitter. Hasil pengujian menggunakan pembobotan fitur jumlah retweet (non-tekstual) menghasilkan akurasi 60% dengan nilai k=3 menggunakan KNN. Penelitian tentang Klasifikasi Topik pada Lirik Lagu dengan Metode Multinomial Naïve Bayes membahas tentang mencari kategori topik dari sebuah lagu. Data yang digunakan berupa lirik lagu yang kemudian akan dilakukan proses *feature extraction* dengan metode *bag of words.* Selanjutnya pembangunan klasifikasi menggunakan *multinomial naive bayes*.Dari penggunaan metode *bag of words* dan klasifikasi *multinomial naive bayes* tersebut menghasilkan nilai perhitungan akurasi sebesar 96% dan perhitungan *f1-measure* sebesar 88.91%. Penelitian lain tentang Analisis Sentimen Pada Ulasan Buku Berbahasa Inggris Menggunakan Information Gain Dan Support Vector Machine membahas tentang mencari sentimen positif dan negatif pada ulasan buku berbahasa inggris. Hasil yang didapat diharapkan dapat memberikan kualitas buku yang lebih baik untuk kedepannya. *Feature selection* yang digunakan adalah TF-IDF dan pada tahap *preprocessing*-nya membandingkan antara tingkat performansi menggunakan *lemmatization* dan tanpa *lemmatization.* Hasil uji membuktikan tanpa *lemmatization* memberikan nilai rata-rata *F1-score* 82.35% lebih besar dibandingkan menggunakan *lemmatization.* Dapat disimpulkan bahwa klasifikasi yang dilakukan tanpa *lemmatization* dan menggunakan *feature selection information gain* menghasilkan *F1-score ­*sebesar 85.34%. Maka penulis ingin membandingkan antara dua klasifikasi yaitu K-Nearest Neighbor dan Multinomial Naive Bayes dengan perbedaan menggunakan stemming dan tanpa stemming pada tahap preprocessing. *Feature extraction* yang digunakan untuk K-Nearest Neighbor adalah tf-idf, sedangkan Multinomial Naive Bayes menggunakan *Bag of Words.*

1. **Metode Penelitian**

**Rancangan Sistem**

Rancangan sistem yang dibuat dalam penelitian ini adalah sebuah perbandingan dua klasifikasi untuk mencari akurasi terbaik dalam menemukan kelas sentimen positf dan negatif pada sebuah data. Data tersebut berupa teks berbahasa indonesia yang berisikan tanggapan mahasiswa terhadap layanan pusat bahasa. Kemudian ditahap *preprocessing* dilakukan penggunaan *stemming* dan tanpa *stemming,* dengan tujuan untuk melihat pengaruh *stemming* terhadap klasifikasi K-Nearest Neighbors dan Multinomial Naive Bayes. Proses dari pembelajaran ini diilustrasikan pada Gambar 1.



Gambar 1. Flowchart system

***Dataset***

Penelitian ini menggunakan dataset teks berbasis Bahasa Indonesia yang berisikan tentang tanggapan mahasiswa terhadap layanan pusat bahasa di Telkom University. Data yang digunakan sebanyak 5000 tanggapan dalam bentuk csv. Terdapat dua kategori pada data tersebut yang harus dicari sentimennya, yaitu kategori tentang fasilitas dan layanan. Sehingga setiap kategori harus menentukan sentiment positif maupun negatif. Data awal diberi label terlebih dahulu secara manual disetiap kalimat tanggapan dengan label 0 (negatif) dan 1 (positif). Untuk pelabelan terhadap setiap kalimat, jika didalam kalimat hanya membahas kekurangan dari kategori layanan saja maka kategori fasilitas bersentimen positif meskipun tidak dibahas dalam kalimat tersebut. Untuk menambah keakuratan perkiraan kelas pada data, maka data *testing* dan *training* dipilih menggunakan *K-Fold Cross Validation* dengan jumlah k = 10 untuk menambah akurasi yang lebih optimal. Berikut ilustrasi data yang digunakan dalam penelitian ini.

TABLE I Ilustrasi Data

|  |  |  |
| --- | --- | --- |
| Kalimat Tanggapan | Label | |
| Fasilitas | Layanan |
| Layanan esapnya mohon segera dibenahi | 1 | 0 |
| Kursinya ditambah lagi | 0 | 1 |
| Perbaiki lagi | 0 | 0 |

***Pre-processing***

*Pre-processing* merupakan tahap dimana data dari dokumen teks yang tidak terstruktur diolah menjadi data terstruktur. *Pre-processing* juga dimanfaatkan untuk Melakukan transformasi atribut dengan mentransformasi atribut ke dalam suatu format data yang lebih mudah untuk meningkatkan efektifitas dan efisiensi yang kemudian di proses dengan teknik klasifikasi[3]. Terdapat beberapa teknik untuk melakukan *pre-processing,* yaitu *Tokenizing, Case folding, Remove punctuation, Stopword removal,* dan *Stemming*. *Tokenizing* merupakan teknik pemotongan teks dari sebuah kalimat menjadi beberapa kata terpisah. *Case folding* merupakan teknik untuk merupah huruf kapital ke dalam bentuk huruf kecil. *Remove punctuation* teknik untuk menghilangkan tanda baca karena tanda baca tidak diperlukan pada kasus ini. *Stopword removal* merupakan tahap untuk membuang kata sambung atau kata yang kurang relevan, contoh kata yang termasuk *stopword* ‘yang’, ‘tapi’, ‘meskipun’. *Stemming* merupakan teknik untuk menghilangkan imbuhan pada sebuah kata, contoh kata ‘meminum’ menjadi ‘minum’. Berikut contoh hasil dari tahap preprocessing.

TABLE II. Preprocessing dengan *stemming*

|  |  |
| --- | --- |
| Kalimat Tanggapan | Hasil Preprocessing |
| Layanan esapnya mohon segera dibenahi | ‘layan’, ‘esap’, ‘mohon’, ‘segera’, ‘benah’ |
| Kursinya ditambah lagi | ‘kursi’, ’tambah’, ’lagi’ |
| Perbaiki lagi | ‘baik’, ‘lagi’ |

TABLE III. Preprocessing tanpa *stemming*

|  |  |
| --- | --- |
| Kalimat Tanggapan | Hasil Preprocessing |
| Layanan esapnya mohon segera dibenahi | ‘layanan’, ‘esapnya’, ‘mohon’, ‘segera’, ‘dibenahi’ |
| Kursinya ditambah lagi | ‘kursinya’, ’ditambah’, ’lagi’ |
| Perbaiki lagi | ‘perbaiki’, ‘lagi’ |

***Feature Extraction***

Feature extraction adalah tahap yang dilakukan setelah tahap *pre-processing* selesai dilakukan. Proses yang akan dilakukan pada tahap ini adalah melakukan penambangan fitur dari hasil *pre-processing* data komentar layanan pusat bahasa. Feature extraction yang digunakan pada penelitian ini adalah *Term Frequency-Inverse Document Frecuency* (TF-IDF) untuk klasifikasi *K-Nearest Neighbor*s, dan *Bag of Words* (BoW) untuk klasifikasi *Multinomial Naïve Bayes*. Inverse Document Frequency (TF-IDF), merupakan sebuah metode ekstraksi fitur untuk menghitung bobot dari sebuah kata. TF-IDF diperlukan untuk menghitung klasifikasi yang menggunakan vector. Bag of Words (BoW), merupakan metode ekstraksi fitur yang digunakan dalam pemrosesan Bahasa alami dan pencarian informasi. BoW diperlukan untuk menghitung klasifikasi yang menggunakan probabilitas[5].

***K-Fold Cross Validation***

*K-Fold Cross Validation* merupakan sebuah teknik untuk membagi data menjadi k himpunan bagian[A]. Penggunaan teknik ini dapat mengurangi waktu komputasi dan mempertahankan estimasi keakuratan. Penentuan jumlah k dapat ditentukan sesuai kebutuhan penelitian yang diperlukan. Pada penelitian ini menggunakan jumlah k=10 yang secara umum banyak digunakan dalam penelitian lain.

***K-Nearest Neighbor***

*K-Nearest Neighbor* merupakan salah satu metode klasifikasi yang proses penentuan label kelas dinyatakan dengan jumlah tetangga terdekat berdasarkan jumlah K yang digunakan. Dari K tetangga terdekat yang terpilih kemudian dilakukan *voting* kelas dari K tetangga terdekat tersebut. Kelas dengan jumlah suara tetangga terbanyaklah yang diberikan sebagai label kelas hasil prediksi pada data uji tersebut. Jumlah K yang digunakan pada KNN dapat mempengaruhi kinerjanya. Sehingga pada penelitian ini jumlah K yang digunakan adalah K=3, K=5, K=7, dan K=9. Dari sejumlah nilai K tersebut, dapat terlihat perbedaan hasil kinerjanya yang paling baik.

Untuk mengukur jarak kedekatan antara data baru dengan data lama salah satunya menggunakan pendekatan *Euclidean distance.* Sebelum menghitung nilai *Euclidean distance,* harus melalui tahapan dalam proses ekstraksi fitur tf-idf. Selanjutnya data hasil dari tf-idf dihitung kemiripannya dengan persamaan sebagai berikut.

*Euclidean Distance :* (1)

adalah nilai tf-idf dari data yang labelnya sudah diketahui, sedangkan adalah nilai tf-idf dari data yang ingin diketahui labelnya. Setelah nilai *euclidean distance* didapatkan, kemudian nilai tersebut diurutkan dari yang paling kecil. Nilai e*uclidean* yang paling kecil merupakan jarak yang paling dekat dengan data yang ingin diketahui labelnya. Sehingga dipilih nilai yang paling dekat sebanyak jumlah K.

TABLE IV. Pencarian label pada KNN

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Data ke - | Nilai hasil TF-IDF | | | | Label Layanan | Label Fasilitas |  |  | Label Layanan | Label Fasilitas |
| Data Training () | 1 | 1.4 | 2 | 0.7 | 0 | 1 | 0 |  | 1.558845727 | 1 | 0 |
| 2 | 1 | 0 | 1.8 | 0.9 | 1 | 1 |  | 2.085665361 | 1 | 1 |
| 3 | 0.5 | 0.3 | 1.1 | 0.6 | 1 | 0 | => | 1.204159458 | 1 | 0 |
| 4 | 0.8 | 1.2 | 1.5 | 0.7 | 0 | 0 |  | 2.004993766 | 0 | 0 |
| Data Testing ( | 5 | 0.1 | 1.3 | 1.2 | 0 | ? | ? |  |  | 1 | 0 |

Tabel IV mengilustrasikan perhitungan KNN menggunakan *euclidean distance*. Pada kolom d(x,y) merupakan hasil dari perhitungan e*uclidean*. Pilih nilai yang paling kecil dengan K=3, kemudian ambil label yang bersesuaian dengan posisi nilai *euclidean* dan pilih yang paling dominan untuk diberikan ke data yang belum berlabel*.*

***Multinomial Naïve Bayes***

Multinomial Naïve Bayes, merupakan metode klasifikasi supervised learning untuk menentukan nilai probabilistik sebuah dokumen terhadap suatu kelas yang berfokus pada text classification. Algoritma ini memiliki ciri utama asumsi yang kuat (naïve) terhadap ketertidakgantungan (independence). Selain itu, algoritma ini memiliki dua asumsi yaitu asumsi *Bag of Words* yang berasumsi bahwa posisi kata dalam dokumen tidak berpengaruh dan Conditional Independence dimana masing-masing probabilitas setiap feature saling independence[1]. Perhitungan Multinomial Naïve Bayes dirumuskan pada persamaan (2)

|  |  |
| --- | --- |
|  | (2) |

Dimana 𝑃(𝑐|𝑑) merupakan posterior probability dokumen terhadap suatu kelas, 𝑃(c) merupakan prior probability suatu kelas, 𝑃(𝑡k|𝑐) merupakan conditional/likelihood probability sebuah term dengan given class c, dan 𝑛d merupakan jumlah term pada dokumen[1]. Posterior probability yang nilainya paling tinggi diambil untuk menentukan kelas hasil prediksi. Formula untuk prior probability dapat dilihat pada persamaan (3).

|  |  |
| --- | --- |
|  | (3) |

Nc merupakan jumlah dari kategori c, sedangkan N merupakan total jumlah kategori. Selanjutnya formula dari *likelihood probability* dapat dilihat pada persamaan (4).

|  |  |
| --- | --- |
|  | (4) |

adalah untuk menghitung jumlah kata pada data training dengan kategori c, kata yang dihitung pada data training mengacu pada kata di data testing. Sedangkan +1 merupakan proses smoothing agar tidak ada nilai 0 pada *likelihood probability.* merupakan jumlah seluruh kata pada kategori c, sedangkan adalah jumlah seluruh kata pada data training tanpa mengambil kembali kata yang sama. Berikut contoh perhitungan *Multinomial Naïve Bayes* dengan menggunakan contoh data training yang sudah melewati tahap *preprocessing* sedangkan data testing tidak dilakukan *preprocessing.*

TABLE V. CONTOH DATASET

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Data ke- | Teks | Label layanan | Label fasilitas |
| Data Training | 1 | ‘layan’, ‘esap’, ‘mohon’, ‘segera’, ‘benah’ | 1 | 0 |
| 2 | ‘kursi’, ’tambah’, ’lagi’ | 0 | 1 |
| 3 | ‘baik’, ‘lagi’ | 0 | 0 |
| Data Testing | 4 | ‘layan’, ‘sudah’, ‘baik’ | ? | ? |

|  |  |  |
| --- | --- | --- |
| Priors probability (layanan) : | P(1) = 1/3, | P(0) = 2/3 |
| Priors probability (fasilitas) : | P(1) = 1/3, | P(0) = 2/3 |

*Likelihood probability* dengan *smoothing*

|  |  |  |  |
| --- | --- | --- | --- |
| *Likelihood probability* (layanan) : | | | *Likelihood probability* (fasilitas) : |
| P(layan | 1) | = |  | |
| P(sudah | 1) | = |  | |
| P(baik | 1) | = |  | |
| P(layan | 0) | = |  | |
| P(sudah | 0) | = |  | |
| P(baik | 0) | = |  | |

1. **Evaluasi**
2. **Kesimpulan**

**Daftar Pustaka**

[1] R. Abdul Aziz and M. Syahrul Mubarok, “Klasifikasi Topik pada Lirik Lagu dengan Metode Multinomial Naïve Bayes,” *Indosc 2016*, no. August, pp. 139–148, 2016.

[2] M. H. Syahnur, M. A. Bijaksana, and M. S. Mubarok, “Kategorisasi Topik Tweet di Kota Jakarta , Bandung , dan Makassar dengan Metode Multinomial Naïve Bayes Classifier Tweet Topic Categorization in Jakarta , Bandung , and Makassar with Multinomial Naïve Bayes Classifier,” vol. 3, no. 2, pp. 3612–3620, 2016.

[3] D. D. Palmer, “Text preprocessing,” *Handbook of*, 2010. [Online]. Available: https://informatikalogi.com/text-preprocessing.

[4] A. D. Saputra, M. S. Mubarok, F. Informatika, and U. Telkom, “KLASIFIKASI SENTIMEN PADA LEVEL ASPEK TERHADAP ULASAN PRODUK BERBAHASA INGGRIS MENGGUNAKAN BAYESIAN NETWORK ( CASE STUDY : DATA ULASAN PRODUK AMAZON ) ASPECT-LEVEL SENTIMENT ANALYSIS ON ENGLISH PRODUCT REVIEWS USING,” vol. 4, no. 3, pp. 4882–4891, 2017.

[5] S. George and S. Joseph, “Text Classification by Augmenting Bag of Words (BOW) Representation with Co-occurrence Feature,” *IOSR J. Comput. Eng.*, vol. 16, no. 1, pp. 2278–8727, 2014.

[6] A. R. Putra, A. Rizal, and M. S. Mubarok, “Klasifikasi Kanker Usus Besar menggunakan Metode GLCM dan JST Backpropagation,” *J. Elektro*, vol. 5, no. 2, pp. 125–132, 2012.

[7] B. A. Nazief and M. Adriani, "Confix Stripping: Approach to Stemming Algorithm for Bahasa Indonesia," Internal publication, Faculty of Computer Science, University of Indonesia, Depok, Jakarta, 1996.

##### Lampiran

Lampiran dapat berupa detil data dan contoh lebih lengkapnya, data-data pendukung, detail hasil pengujian, analisis hasil pengujian, detail hasil survey, surat pernyataan dari tempat studi kasus, screenshot tampilan sistem, hasil kuesioner dan lain-lain.