

BAB I

PENDAHULUAN

1.1 Latar Belakang

Layanan akademik memegang peranan yang sangat penting bagi civitas akademik, khususnya bagi mahasiswa. Sistem layanan akademik menjadi suatu media bagi universitas untuk menyampaikan informasi. Sehingga dengan adanya layanan akademik mahasiswa dapat dengan mudah dalam mencari informasi yang berkaitan dengan akademik, jadwal, mata kuliah dan lain sebagainya, yang berkaitan dengan informasi seputar perkuliahan. Namun, sebagian besar dalam penyampaian informasi ini, layanan akademik masih dilakukan oleh *customer service* yang tidak memiliki waktu banyak dalam memberikan berbagai informasi dan melayani mahasiswa selama 24 jam, hal ini karena keterbatasan waktu kerja yang telah ditetapkan dan hanya berpaku selama jam kantor berlangsung. Sehingga dampaknya adalah informasi yang diterima oleh mahasiswa akan lambat dan pemanfaatan waktu yang tidak efisien.

Adapun layanan akademik yang tersedia saat ini di UIN Sunan Gunung Djati Bandung salah satunya adalah website sapa.uinsgd.ac.id. Namun layanan akademik tersebut terbatas karena hanya bergantung pada jam kantor.

Penyampaian informasi pada web tersebut juga kurang interaktif dan tidak memenuhi keinginan mahasiswa karena tidak ada fitur tanya jawab secara langsung pada web tersebut, serta keterbatasan customer service yang tidak bisa melayani 24 jam penuh pengunjung web tersebut.

Menurut Muhammad Fakhri, salah satu mahasiswa jurusan Teknik Informatika 2016, mengutarakan pengalaman kesulitannya dalam menjangkau informasi terkait layanan informasi, misal yang dialami salah satunya adalah pendaftaran sidang, untuk bertanya pada *customer service* atau salah satu admin yang mengetahui informasi tersebut tidak fast respon dalam menjawab pertanyaannya, sedangkan dia juga memahami bahwa waktu yang dia gunakan untuk bertanya di luar jam kerja. Sehingga dia harus menunggu waktu yang pas untuk kembali menanyakan hal tersebut.

Untuk itu, dapat dibuatkan sebuah aplikasi yang dapat digunakan sebagai pengganti *customer service* berupa sistem aplikasi chatbot. Chatbot adalah sebuah program komputer yang dirancang untuk mensimulasikan sebuah tanya jawab atau komunikasi yang interaktif layaknya model diskusi, yang dapat menggunakan bahasa sehari-hari kepada pengguna (manusia) melalui bentuk teks, suara, dan visual [1].

Secara umum, teknik bertanya setiap mahasiswa pasti berbeda-beda, dan chatbot tidak akan langsung dapat memahami pertanyaan yang diajukan

mahasiswa tersebut karena chatbot tidak dapat memahami bahasa alami manusia. Maka sebab itu, diperlukan pendekatan *Natural Language Processing* (NLP) yang bertujuan memproses pertanyaan yang diajukan mahasiswa tersebut agar dapat dipahami oleh chatbot. *Natural Language Processing* merupakan *artificial Intelligence* (kecerdasan buatan) yang membuat mesin mampu menjawab pertanyaan secara cerdas sekarang ini dalam bidang rekayasa yang sangat luas. Machine Learning memberi kita teknik terbaik untuk kecerdasan buatan seperti klasifikasi, regresi, pembelajaran terawasi dan pembelajaran tanpa pengawasan dan banyak lagi. Kita dapat menggunakan pengklasifikasi apapun seperti Decision Tree, Naïve Bayes, dan lain sebagainya[2].

Pada penelitian sebelumnya, pada jurnal yang berjudul Perbandingan Klasifikasi Rocchio Dan *Multinomial Naïve Bayes* Pada Analisis Sentimen Data Twitter Bahasa Indonesia oleh Pantouw, bahwa hasil akurasi yang dihasilkan menggunakan metode *Multinomial Naïve Bayes* rata-rata sebesar 72.193%. hal tersebut menunjukkan akurasi yang lebih tinggi apabila dibandingkan dengan metode Rocchio dengan nilai 70.673%. Kemudian dilakukan proses pelabelan ulang sentimen dengan bantuan clustering. Setelah pelabelan ulang hasil klasifikasi ditemukan sebesar 85.399% untuk Mulltinomial Naïve Bayes dan sebesar 96.283% untuk Rocchio dengan waktu komputasi Rocchio lebih cepat sebesar 35.70%.

Adapun pada penelitian yang berjudul Automatic Summarization Text On Multi Document Using *Textrank* oleh Putra dan Dewi Evita dalam studi sebelumnya onsummary single document menggunakan *Textrank* memperoleh akurasi 95% dan baik dalam meringkas pesan teks, sehingga dalam studi ini akan diukur akurasi *Textrank* pada ringkasan multi dokumen. Dari hasil tes yang dilakukan, menunjukkan akurasi 74%, penarikan kembali 73%, precision sebesar 74%, dan fmeasure sebesar 73%. Oleh karena itu peneliti menyimpulkan algoritma *Textrank* cocok diterapkan untuk penelitian sebagai *keyword* extraction pada chatbot [3].

Kemudian pada penelitian oleh Dea, Herwinda dkk (2017), yang berjudul “Implementasi Algoritma Multinomial Naive Bayes Classifier Pada Sistem Klasifikasi Surat Keluar” bertujuan untuk menghasilkan pencarian data dengan waktu yang efisien melalui tahap klasifikasi. Peneliti melakukan analisa terhadap Multinomial Naive Bayes Classifier dalam mengklasifikasi surat keluar sehingga dapat menentukan nomor surat secara otomatis. Sistem klasifikasi didukung dengan confix-stripping stemmer untuk menemukan kata dasar dan untuk pembobotan kata menggunakan TF-IDF. Dari hasil pengujian menunjukkan bahwa implementasi Multinomial Naive Bayes Classifier pada sistem klasifikasi surat memiliki tingkat accuracy, *Precision*, Recall, dan F-measure berturut-turut sebesar 89.58%, 79.17%, 78.72%, dan 77.05% [4].

Dari jurnal di atas yang dijadikan acuan untuk menggunakan metode klasifikasi yaitu algoritma *Multinomial Naïve Bayes* dan *Textrank* yang nantinya akan digunakan peneliti untuk membuat chatbot yang bisa digunakan dalam menanggapi pertanyaan masyarakat dengan memberikan jawaban secara interaktif dan otomatis.

Multinomial Naive Bayes Classifier merupakan model pengembangan dari algoritma bayes yang cocok dalam pengklasifikasian jawaban yang diajukan user pada aplikasi chatbot yang akan dibangun. Pada formula Multinomial Naive Bayes Classifier, kelas jawaban tidak hanya ditentukan dengan kata yang muncul tetapi juga jumlah kemunculannya. Algoritma ini disebut “*naive*” karena mengasumsikan independensi di antara kemunculan kata-kata di jawaban yang datanya nanti disimpan sebagai knowledge base, tanpa memperhitungkan urutan kata dan informasi konteks dalam kalimat atau data jawaban yang telah disediakan.

Selain Algoritma MNB, pada chatbot ini diterapkan juga algoritma Textrank untuk mengidentifikasi entity dengan mengekstrak kata kunci. Dengan demikian, hasil dari ekstraksi kata kunci tersebut akan digunakan dalam pencarian jawaban dengan mencocokkan hasil keywords yang didapatkan menggunakan perbandingan SequenceMatcher[5].

Berdasarkan pemaparan di atas, dapat dilakukan penelitian "**Implementasi Algoritma Multinomial Naive Bayes dan *Textrank* Untuk Chatbot Layanan Akademik**". Dalam penelitian ini diharapkan dapat membantu civitas akademik khususnya mahasiswa untuk memudahkan dalam mendapatkan informasi dengan waktu yang lebih efisien dan informasi layanan akademik akan didapatkan secara relevan dan jelas karena didapat berdasarkan sumber informasi yang sama.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang dijelaskan di atas, peneliti menyusun beberapa rumusan masalah terkait dalam permasalahan tersebut sebagai berikut :

1. Bagaimana cara menerapkan *Algoritma Multinomial Naive Bayes* sebagai *intent classifier* pada sistem *chatbot* layanan akademik?
2. Bagaimana menerapkan metode *Textrank* sebagai *keyword extraction* pada *chatbot* layanan akademik?
3. Bagaimana hasil akurasi jawaban yang didapatkan pada sistem *Chatbot* layanan akademik?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, terdapat beberapa tujuan terkait dalam permasalahan tersebut yaitu sebagai berikut.

1. Menerapkan Algoritma *Multinomial Naive Bayes* untuk mengklasifikasi pertanyaan pada chatbot.
2. Menerapkan algoritma *Textrank* sebagai *keyword extraction* pada chatbot.
3. Mengetahui tingkat akurasi jawaban yang didapatkan pada sistem Chatbot layanan akademik.

1.4 Batasan Masalah

Berdasarkan rumusan masalah yang telah dijelaskan di atas, peneliti membatasi masalah yang akan dianalisa dalam pembuatan sistem ini. Adapun batasan-batasan tersebut adalah sebagai berikut.

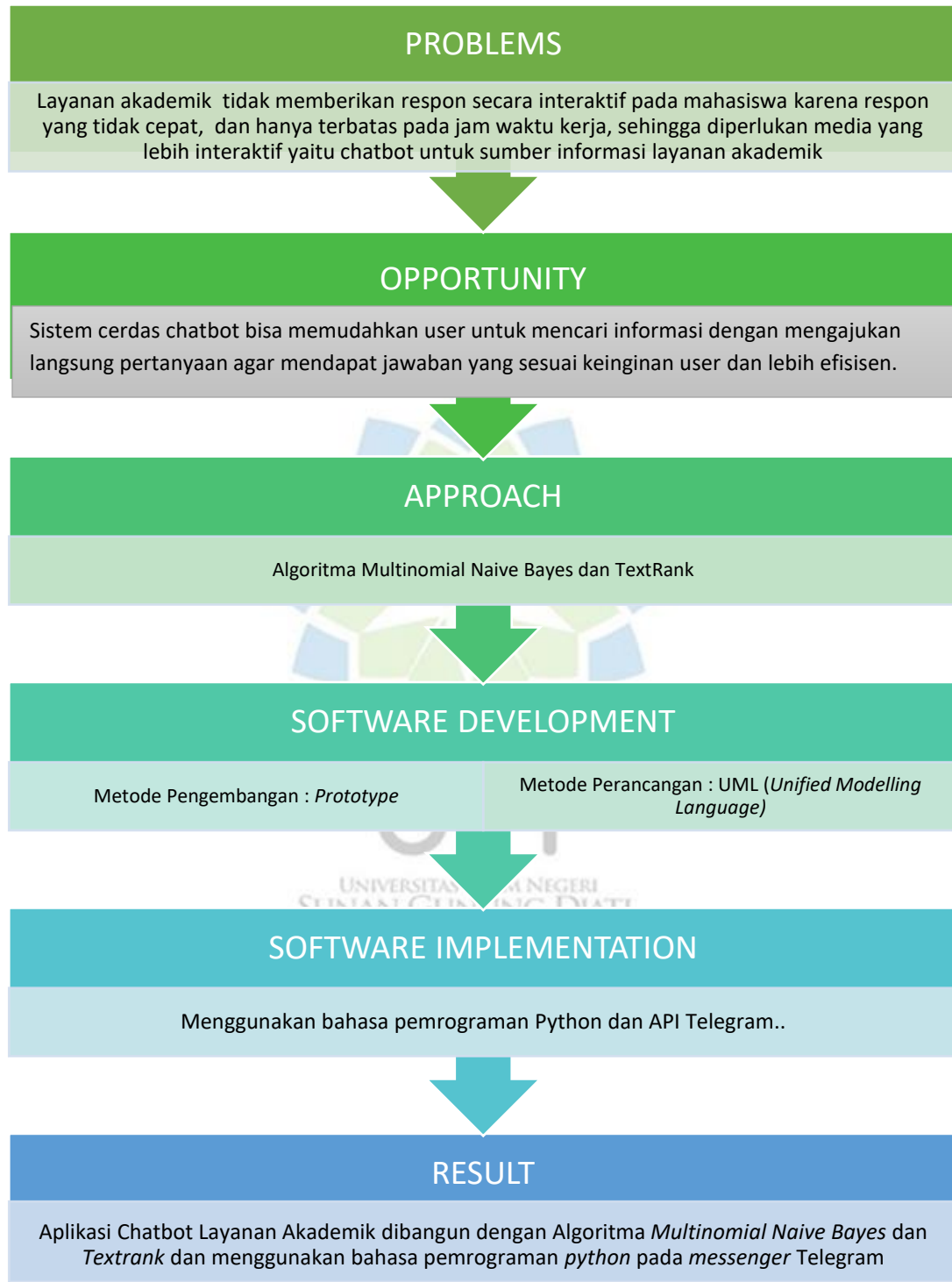
1. Data yang digunakan merupakan data informasi layanan akademik yang didapat dari berbagai sumber layanan akademik seperti website simak, atau wawancara kepada civitas akademik yang mempunyai pengetahuan dan wewenang dalam mengelola sumber informasi layanan akademik UIN Sunan Gunung Djati Bandung;.
2. Penempatan Algoritma Multinomial Naive Bayes digunakan untuk mengklasifikasikan pertanyaan;
3. Penempatan Algoritma Textrank digunakan sebagai keyword extraction;
4. Aplikasi chatbot layanan akademik adalah aplikasi chatbot menggunakan API Telegram;
5. Percakapan hanya dapat menggunakan Bahasa Indonesia;

6. Aplikasi hanya bisa merespon percakapan dalam bahasa alami dan tidak dapat melakukan perhitungan matematis;
7. Aplikasi tidak dapat membedakan lawan bicara berdasarkan jenis kelamin, umur atau nama;
8. Aplikasi hanya dapat menjawab pertanyaan yang mengandung *keywords* tentang layanan akademik UIN Sunan Gunung Djati Bandung di Fakultas Sains dan Teknologi Jurusan Teknik Informatika;
9. Dataset yang digunakan adalah data akademik tahun ajaran 2020/2021
10. Aplikasi hanya dapat menjawab pertanyaan dengan benar jika jumlah minimal empat kata sesuai dengan dataset
11. Aplikasi yang dibangun menggunakan bahasa *Python*;
12. Aplikasi hanya akan menjawab pertanyaan berkaitan dengan label layanan akademik jurusan Teknik Informatika sebagai berikut.
 - a. Sidang
 - b. Jadwal
 - c. Dosen
 - d. Mata kuliah
 - e. Jurusan

1.5 Kerangka Pemikiran

Kerangka pemikiran merupakan penjelasan secara general dari alur logika berjalannya sebuah penelitian. Berikut ini merupakan kerangka penelitian dari tugas akhir ini.





Gambar 1. 1 Kerangka Pemikiran

1. 6 Pengujian

Pengujian adalah proses yang bertujuan untuk menemukan kesalahan dalam sistem. Pada penelitian ini dilakukan pengujian aplikasi / service dan pengujian pada algoritma yang diterapkan yaitu Multinomial Naïve Bayes dan Textrank.

1. Pengujian Aplikasi / Service

Pada pengujian aplikasi / service, jika ada kemungkinan pengujian yang dilakukan akan menemukan error yang tidak ditampilkan, maka kondisi sistem dinyatakan baik. Jika pengujian dapat mendeteksi kesalahan yang awalnya tidak ditemukan, pengujian dinyatakan berhasil. Di antara jenis pengujian yang ada, pengujian *black box* adalah salah satunya. Pengertian pengujian kotak hitam adalah proses pengujian yang mengabaikan mekanisme internal komponen atau sistem dan berfokus pada kondisi eksekusi dan nilai keluaran hasil sebagai respons terhadap nilai masukan yang dipilih.

Untuk mengidentifikasi kesalahan pada sistem yang dibuat, teknik pengujian yang digunakan adalah pengujian *black box*. Pengujian *black box* berfokus pada spesifikasi fungsional perangkat lunak. Penguji dapat melakukan proses identifikasi dari kumpulan kondisi input dan menguji spesifikasi fungsional sistem.

Black box Testing lebih dominan menemukan hal-hal seperti di bawah ini [9].

1. Fungsi yang tidak ada atau salah.
2. Kesalahan pada antarmuka (*interface errors*).
3. Akses database dan kesalahan struktur data.
4. Kesalahan pada performansi atau kinerja
5. Kesalahan inisialisasi dan terminasi.

Testing dirancang untuk dapat menjawab pertanyaan-pertanyaan berikut[9].

1. Bagaimana suatu fungsi dapat diuji agar dapat dikatakan valid?
 2. Masukan apa yang merupakan bahan kasus uji yang baik?
 3. Apakah sistem sensitif terhadap *input* tertentu?
 4. Bagaimana cara mengisolasi kumpulan data?
2. Pengujian Model Multinomial Naïve Bayes dan Textrank

Pada pengujian algoritma MNB dan Textrank dalam satu kesatuan aplikasi digunakan pengujian Confusion Matrix. Yaitu pengujian untuk mengukur performa dari klasifikasi yang dihasilkan. Cara kerjanya yaitu membandingkan nilai aktual dengan nilai prediksi yang dari partisi data yang dihasilkan oleh klasifikasi tersebut.

1.7 Metodologi

Adapun metodologi yang digunakan dalam penelitian ini adalah sebagai berikut :

1.7.1 Metode Pengumpulan Data

Metode pengumpulan data pada penelitian ini terdiri dari 3 tahap, yaitu:

1. Observasi

Melakukan observasi dengan peninjauan serta penelitian langsung terhadap suatu permasalahan yang di angkat. Tepatnya pada tanggal 26 Oktober 2020, peneliti melakukan observasi pada salah satu pihak admin yang ada di Jurusan untuk izin menganalisa beberapa data yang nantinya akan dijadikan sebagai sumber *dataset* sehingga data tersebut dapat dijadikan sebagai informasi yang relevan karena didapatkan dari pihak yang berwenang dan mengetahui langsung tentang data perkuliahan di jurusan.

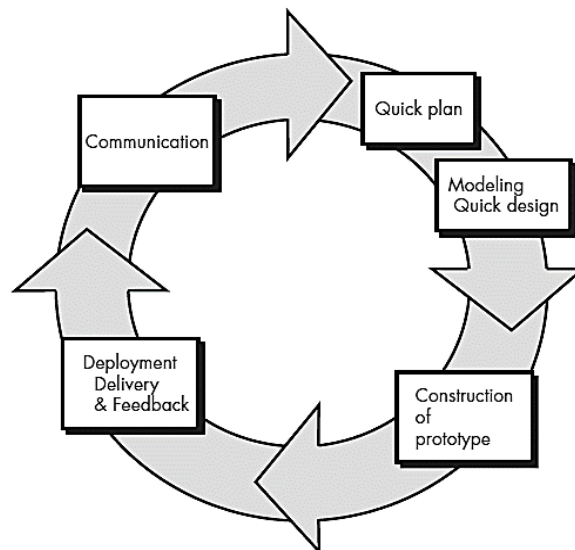
2. Studi Literatur

Studi Literatur yang dilakukan adalah dengan mengumpulkan dan mempelajari berbagai tinjauan pustaka, jurnal, serta artikel yang berkaitan dengan sistem yang akan dibangun. Salah satu tinjauan pustaka yang dijadikan acuan adalah jurnal yang berjudul “Otomatisasi Layanan *Frequently Ask Questions* Berbasis *Natural Language Processing* Pada Telegram Bot” sehingga peneliti mengusulkan sebuah solusi untuk memaksimalkan layanan informasi akademik yang berbasis *Natural Language Processing* (NLP). Layanan ini dilakukan dengan fitur bot pada messenger Telegram menggunakan API yang sudah tersedia dan dirancang berbasis NLP yang dibangun menggunakan bahasa pemrograman *python*. Dengan begitu chatbot ini bertindak

sebagai *customer service* yang akan menjawab pertanyaan-pertanyaan yang diinputkan oleh visitor.

1.7.2 Metode Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak yang digunakan pada penelitian ini yaitu metode pengembangan perangkat lunak *Prototype*. Metode ini dapat memudahkan dalam proses membangun sebuah sistem karena *developer* dan *client* dapat berinteraksi selama proses pembangunan sistem. Seperti yang digambarkan pada Gambar 2.1 dibawah ini.



Gambar 2. 1 Metode Pengembangan Perangkat Lunak Prototype

Adapun tahapan dalam metode pengembangan perangkat lunak prototype berdasarkan sumber Media Jurnal Informatika, dengan judul “Rancang Bangun Digital Archiving Di Perguruan Tinggi Menggunakan Metode Prototype Model Studi Kasus: Universitas Suryakencana” Syaddad, Hasbu Naim (2017) [6] yaitu sebagai berikut.

1. *Communication*

Developer dengan *Client* melakukan interaksi dengan bertemu secara langsung untuk menentukan suatu tujuan, seperti *requirements* yang diinginkan dan gambaran bagian-bagian apa saja yang akan dibutuhkan.

2. *Quick Plan*

Perancangan akan diproses secara cepat dan mewakili semua aspek *software* yang diketahui, dan rancangan ini menjadi landasan pembangunan *prototype*.

3. *Modelling Quick Design*

Fokus pada representasi aspek perangkat lunak yang dapat dilihat oleh *client*. *Modelling Quick Design* ini lebih kepada ke pembuatan *prototype*.

4. *Construction of Prototype*

Mengkodekan sebuah sistem berdasarkan rancangan *prototype* dari *software* yang akan dibangun.

5. *Deployment Delivery & Feedback*

Prototype yang sudah berhasil dibuat oleh *developer* kemudian disampaikan kepada *client*, untuk dilakukan dievaluasi, dan *client* akan memberikan *feed back* dan memberikan revisi *software requirements*. Proses ini akan terus berulang sampai semua *requirements* terpenuhi.

1.8 Sistematika Penulisan

Tugas akhir ini dibagi ke dalam 5 (lima) BAB, dengan susunan sistematika penulisan di bawah ini :

BAB I PENDAHULUAN

Pada BAB 1, terdapat tujuh pokok bahasan, diantaranya adalah latar belakang suatu masalah, perumusan masalah, tujuan, batasan-batasan masalah dan manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada BAB II, berisi sebuah teori yang menjadi dasar dan landasan untuk digunakan sebagai tinjauan untuk memecahkan masalah dan mendukung penelitian ini.

BAB III ANALISIS DAN PERANCANGAN

Pada BAB III, berisi mengenai kebutuhan sistem yang telah dianalisis sebelumnya, kemudian data dan algoritma yang menjadi landasan untuk perancangan sistem. Bab ini juga berisi sebuah arsitektur sistem yang akan dibuat.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Pada BAB IV, menerangkan detail hasil implementasi dan pengujian aplikasi yang sudah dibangun.

BAB V PENUTUP

Pada BAB V, berisi kesimpulan dari penelitian yang dilakukan dan mewakili isi secara keseluruhan. Bab ini juga berisi saran untuk pengembangan penelitian yang akan dilakukan ke depannya.

DAFTAR PUSTAKA

Daftar Pustaka berisi sumber-sumber dari media cetak atau tertulis yang digunakan dalam penelitian kemudian dikutip dalam penyusunan.

LAMPIRAN

Pada lampiran ini berisi proses penyusunan serta perancangan seperti source code serta kelengkapan dokumen lainnya.

