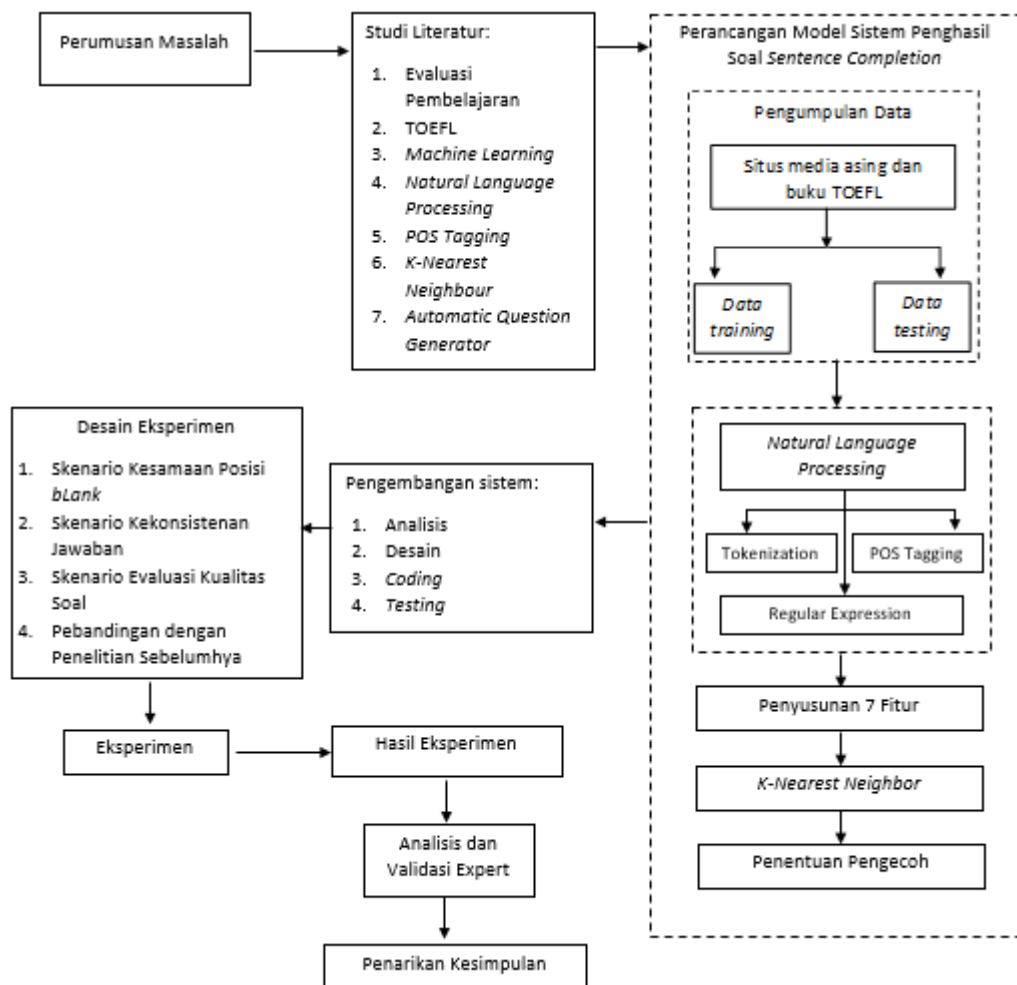


BAB III METODE PENELITIAN

Pada bab ini akan dijelaskan mengenai metodologi penelitian, mulai dari desain penelitian, alat dan penelitian, dan metode penelitian.

1.1 Desain Penelitian

Desain penelitian merupakan kerangka atau gambaran proses kerja yang digunakan untuk melakukan penelitian. Desain penelitian ini digambarkan pada Gambar 3.1 dan akan dijelaskan pada subbab selanjutnya.



Gambar 3. 1 Metode Penelitian

1.1.1. Perumusan Masalah

Tahapan pertama dalam penelitian ini adalah perumusan masalah. Pada tahap ini penulis mengumpulkan latar belakang dan merumuskan masalah apa yang akan dicari solusinya.

Perumusan masalah ini yang akan menjadi panduan bagi penulis dalam melakukan penelitian. Hal ini juga yang mendasari pencarian referensi yang harus dipelajari. Latar belakang dan perumusan masalah akan menghasilkan tujuan dan juga manfaat penelitian yang menjawab rumusan masalah.

1.1.2. Studi Literatur

Studi literatur merupakan tahap yang dilakukan untuk mempelajari dan memahami teori yang akan digunakan untuk penelitian dan penyelesaian masalah. Teori yang harus dipelajari untuk mendukung penelitian ini yaitu mengenai Evaluasi Pembelajaran, TOEFL dan soal Bahasa Inggris, *Machine Learning*, *Natural Language Processing*, *POS Tagging*, algoritma *K-Nearest Neighbor*, dan Penelitian Terkait *Automatic Question Generator*

Pada tahap ini, peneliti mempelajarinya dengan cara membaca dan memahami beberapa literatur seperti buku, jurnal, *paper*, serta artikel yang memuat topik serupa dengan teori-teori yang telah disebutkan.

1.1.3. Perancangan Model Sistem Penghasil Soal *Sentence Completion*

Pada tahap ini, model perancangan sistem penghasil soal *sentence completion* dibuat oleh penulis. Model ini dibangun dengan didasarkan oleh beberapa tahapan penting yaitu:

1. Pengumpulan data

Tahap ini merupakan tahap untuk mengumpulkan koleksi *data training* dan *data testing*. Data-data ini kemudian akan menjadi data penting yang diolah oleh sistem. *Data training* didapatkan dari buku-buku TOEFL dengan tipe soal *sentence completion*. Sedangkan, *data testing* data-datanya diambil dari situs media asing yang terpercaya mulai dari berita hingga penggunaan tata bahasanya.

2. *Natural Language Processing*

Natural Language Processing (NLP) merupakan salah satu bagian dalam kecerdasan buatan yang dikhususkan dalam pemrosesan teks. NLP memiliki banyak teknik yang dapat membantu komputer dalam memahami bagian-bagian teks.

Dalam penelitian ini, pemrosesan yang dibutuhkan yaitu untuk menghapus tanda baca, membagi kalimat menjadi kata, dan juga mengetahui label *part-of-speech* dari suatu kata. Maka, teknik-teknik yang dilakukan adalah *tokenization*, *regular expression*, dan *POS Tagging*.

Tokenization merupakan salah satu teknik NLP yang berfungsi untuk memecah kalimat menjadi kata. Sedangkan, *regular expression* atau biasa disebut *regex* merupakan teknik untuk mencari tanda tertentu dalam suatu kalimat. *Regex* dapat digunakan untuk menghapuskan tanda baca yang diinginkan. Sementara itu, untuk mendapatkan label *part-of-speech*, maka digunakan tahap *POS Tagging*.

3. Penyusunan 7 fitur

Penyusunan 7 fitur merupakan tahap untuk mendapatkan atribut untuk digunakan dalam klasifikasi target kata dalam menentukan posisi *blank*. 7 fitur ini didapatkan dengan beberapa cara yang diantaranya merupakan teknik-teknik NLP yang telah disebutkan. 7 fitur ini digunakan baik pada *data training* maupun *data testing*.

Pada *data training*, 7 fitur ini disusun beserta targetnya. Target bernilai *true* jika kata tersebut merupakan posisi *blank* dalam kalimat, sedangkan *false* diberikan jika kata tersebut bukan merupakan posisi *blank*. Namun, pada *data testing*, fitur tidak disertai target karena target akan ditentukan dalam proses klasifikasi.

4. *k-Nearest Neighbor*

K-Nearest Neighbor (KNN) merupakan salah satu algoritma dalam *machine learning*. KNN sendiri merupakan algoritma dari tipe pembelajaran *supervised learning* dalam bidang klasifikasi. Maka dari itu, KNN menjadi salah satu algoritma yang biasa digunakan dalam klasifikasi teks. Algoritma

klasifikasi tentu saja membutuhkan *data training* yang menjadi proses pembelajaran dalam mengkategorikan suatu *data testing* nantinya.

Pada penelitian ini KNN merupakan tahap klasifikasi untuk menentukan target posisi *blank*. Proses untuk menentukan target atau klasifikasi dalam KNN ditentukan dengan dominan target dari k jarak terdekat. Atribut yang digunakan untuk menghitung jarak adalah 7 fitur yang telah dijelaskan sebelumnya. Target dari proses klasifikasi ini hanya ada 2 macam, yaitu *true* apabila kata tersebut menjadi posisi *blank* atau *false* jika tidak bisa menjadi posisi *blank*.

5. Penentuan pengecoh

Soal yang dihasilkan dalam penelitian ini yaitu soal dengan tipe *sentence completion*. Dalam TOEFL, soal tipe *sentence completion* merupakan soal *fill-in-the-blank* dengan 4 pilihan. Pilihan tersebut terdiri dari 1 jawaban benar dan 3 jawaban salah.

Tentunya, 3 jawaban salah harus merupakan kata sepadan sehingga dapat mengecoh peserta yang mengerjakan soal tersebut. Meskipun kata sepadan, kata yang dijadikan pilihan tidak boleh ambigu agar tidak menghasilkan 2 jawaban benar.

Untuk mencerdaskan sistem dalam menentukan pengecoh, penelitian ini akan membuat aturan heuristik yang digunakan untuk menentukan pengecoh agar soal menjadi berkualitas. Tidak hanya berkualitas dalam pemilihan posisi *blank*, namun juga dalam membuat pengecoh. Selain aturan heuristik, akan digunakan juga kamus *verb* dalam penentuan pengecoh untuk kata kerja yang sangat variatif.

Poin-poin di atas merupakan penjelasan singkat mengenai perancangan model. Selanjutnya, perancangan model akan dibahas lebih detail pada Bab 4.

1.1.4. Pengembangan Sistem

Tahap ini merupakan tahap untuk pembuatan perangkat lunak (*software*) untuk mengolah data yang telah dikumpulkan. Rekayasa perangkat lunak dilakukan dalam beberapa tahap sesuai dengan metode pengembangan perangkat lunak yang dipilih yaitu *waterfall*. Tahap pertama dalam

pengembangan sistem yaitu analisis, pada tahap ini akan dianalisis bagaimana perangkat lunak akan dibuat.

Kemudian, tahap kedua adalah desain. Pada tahap ini akan dibuat desain aplikasi, mulai dari desain *interface*, desain *database*, dan lain sebagainya. Kemudian, masuk ke dalam tahap *coding*, di mana pada tahap ini mulai dilakukan implementasi dari analisis dan desain yang telah dilakukan sebelumnya. Setelah itu, tahap terakhir adalah tahap *testing* atau pengujian aplikasi perangkat lunak yang telah dibuat.

1.1.5. Desain Eksperimen

Tahap ini merupakan tahap yang digunakan untuk mencoba penggunaan metode dengan menggunakan data yang telah dipersiapkan. Pertama dengan cara pemilihan kalimat, lalu dilabeli dengan *POS Tagging*, setelah itu dibuatlah opsi beserta pengecohnya.

1.1.6. Eksperimen

Eksperimen dilakukan sesuai dengan rancangan desain yang telah ditentukan. Eksperimen melibatkan data yang sebelumnya telah dikumpulkan, yaitu artikel berita serta soal-soal TOEFL. Eksperimen juga dilakukan sesuai dengan rancangan skenario.

1.1.7. Analisis dan Validasi *Expert*

Setelah perangkat lunak telah dibuat, tahap selanjutnya analisis dan validasi hasil oleh *expert*. Pada tahap ini perangkat lunak diuji coba sesuai dengan kegunaannya. Soal yang dihasilkan akan diuji oleh *expert* yang kemudian dianalisa oleh penulis untuk menentukan kualitas soal yang dihasilkan.

1.2 Alat dan Bahan Penelitian

Pada subbab ini akan dijelaskan mengenai alat dan bahan penelitian yang digunakan oleh penulis.

1.2.1 Alat Penelitian

Penelitian ini menggunakan alat-alat yang menunjang penelitian. Alat-alat penelitian ini dibagi menjadi 2 aspek, yaitu perangkat keras dan perangkat lunak seperti dijelaskan sebagai berikut:

1. Perangkat Keras (*Hardware*) yaitu sebuah laptop dengan spesifikasi:
 - *Processor* Intel Core i3-3217U
 - *Random Access Memory* (RAM) 4 GB
 - *VGA* Radeon 8530M 1 GB
2. Perangkat Lunak (*Software*) sebagai berikut:
 - XAMPP
 - Microsoft Excel 2016
 - *Web Browser* Google Chrome
 - Sublime Text 2
 - *Command prompt*
 - *Framework* CodeIgniter

1.2.2 Bahan Penelitian

Bahan yang digunakan dalam penelitian ini adalah berita dari media asing yang telah di data. Kemudian, bahan lainnya adalah soal-soal TOEFL yang digunakan sebagai data soal untuk melatih algoritma.

1.3 Metode Penelitian

Metode yang digunakan dalam penelitian ini terbagi menjadi dua bagian, yaitu metode pengumpulan data dan metode pengembangan perangkat lunak.

3.3.1 Metode Pengumpulan Data

Metode pengumpulan yang digunakan penulis untuk mendapatkan data yang akurat demi menunjang penelitian terbagi mejadi tiga cara, yaitu:

1. Studi Literatur

Studi literatur digunakan untuk mempelajari teori-teori yang menunjang penelitian. Teori-teori yang terkait dengan penelitian ini diantaranya yaitu mengenai Evaluasi Pembelajaran, TOEFL dan soal Bahasa Inggris, *Machine Learning*, *Natural Language Processing*, *Information Extraction*, algoritma *K-Nearest Neighbor*, dan *Automatic Question Generator*.

2. Mengumpulkan data soal TOEFL

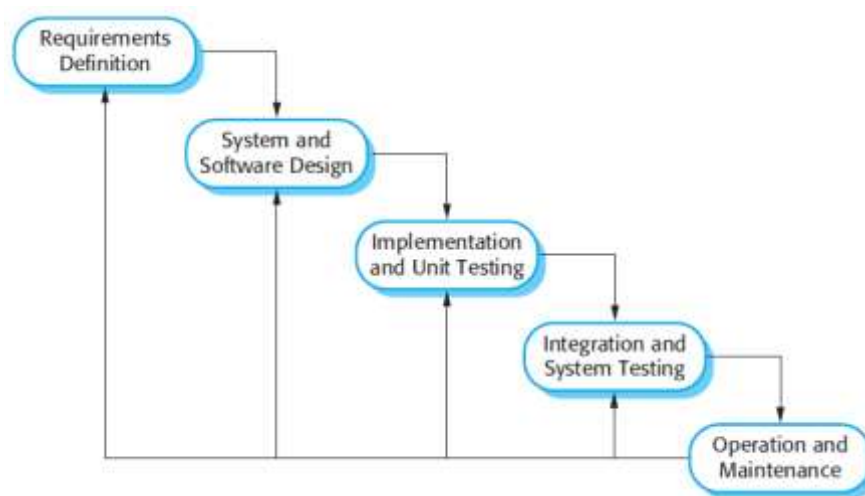
Data soal TOEFL dikumpulkan secara manual yang kemudian dimasukkan ke dalam basis data. Jenis soal TOEFL yang dikumpulkan ialah soal-soal dengan tipe *sentence completion*.

3. Mengunduh corpus berita

Berita dikumpulkan secara otomatis oleh program dengan bahasa *python*. Program yang digunakan merupakan program *library article scrapper* di mana keluarannya adalah sebuah file teks yang berisi paragraf-paragraf berita.

3.3.2 Metode Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak memiliki banyak tipe. Namun, metode yang digunakan penulis dalam mengembangkan perangkat lunak yaitu adalah metode air terjun atau *waterfall*. Model ini sering juga disebut model *sequential linier*. Tahapan model *waterfall* dapat dilihat pada Gambar 3.2 berikut:



Gambar 3. 2 Model Waterfall (Sommerville, 2011)

Pada Gambar 3.2 diperlihatkan bahwa dalam model *waterfall* terdapat lima tahapan. Tahap-tahap ini dapat membantu penyelesaian dalam proses pengerjaan perangkat lunak. Jika suatu waktu terdapat kesalahan, bisa mengacu pada tahap sebelumnya untuk perbaikan. Pengertian dari setiap tahapan dari Gambar 3.2 tersebut dijelaskan oleh Sommerville pada tahun 2011 sebagai berikut:

1. *Requirements Definition (Analysis)*

Pada tahap analisis ini, aplikasi yang akan dibuat dikonsepsi sedemikian rupa. Pada tahapan analisis ini fitur-fitur yang akan dibuat dikonsultasikan juga dengan pengguna sistem. Hal ini akan ditetapkan sebagai spesifikasi sistem yang rinci.

2. *System and Software Design*

Dalam tahapan ini, *software* akan didesain mulai dari basis datanya, arsitektur, hingga *user interface*. Desain sistem dibuat berdasarkan spesifikasi yang telah ditetapkan.

3. *Implementation and Unit Testing*

Pada tahap implementasi ini bisa juga disebut tahap *coding*. Dalam tahap ini, desain perangkat lunak dibentuk sebagai satu program. Program ini kemudian akan diuji untuk memenuhi kebutuhan spesifikasi.

4. *Integration and System Testing*

Tahapan pengujian atau *testing* ini dilakukan untuk setiap bagian dalam program, apakah sudah sesuai kebutuhan spesifikasi atau belum. Selain itu, dalam tahap ini juga setiap bagian program diuji apakah sudah terintegrasi dengan baik atau belum. Jika secara keseluruhan sistem sudah terintegrasi dengan baik dan juga sesuai spesifikasi dan syarat, maka program atau sistem akan dikirim ke pengguna sistem.

5. *Operation and Maintenance*

Dalam tahap ini, sistem mulai diaplikasikan dan juga digunakan. Bisa juga pada tahap ini ditemukan beberapa *error* yang harus diperbaiki dan tidak ditemukan saat pembuatan. Pada tahap ini pun, aplikasi bisa ditambah dengan fitur baru yang sekiranya dibutuhkan pengguna.