BAB I
PENDAHULUAN

I.1 Latar Belakang Masalah

Lembaga Ilmu Pengetahuan Indonesia(LIPI) merupakan salah satu lembaga yang bergerak dalam pengembangan ilmu pengetahuan. Pusat Penelitian Informatika merupakan salah satu lembaga yang berada dibawah koordinasi Lembaga Ilmu Pengetahuan Indonesia(LIPI). Pusat penelitian ini bertujuan untuk penggunaan teknologi dalam pengumpulan, pengolahan, penyimpanan dan penyebarluasan data atau informasi kepada masyarakat dengan cepat, tepat waktu dan akurat. Sehingga akan terjadi percepatan peningkatan kemampuan masyarakat dalam pembangunan yang berkelanjutan.

Berdasarkan hasil wawancara dengan pihak Pusat Penelitian Informatika – LIPI. Menyadari semakin pesatnya perkembangan teknologi media penyimpanan digital telah mendorong terjadinya ledakan jumlah dokumen elektronik yang tersimpan dalam *repository* di lembaga tersebut. Berbagai artikel ilmiah dari berbagai peneliti dan kalangan masyarakat telah tersedia dalam versi digital. Namun, fenomena ini tidak disertai dengan pertumbuhan jumlah informasi atau pengetahuan yang dapat disarikan dari dokumen-dokumen elektronik tersebut, karena sistem yang berjalan sekarang ini hanya sebatas penyimpanan digital dan semua dokumen artikel ilmiah digabung dalam satu kelompok. Hal ini menjadi penghambat dalam mengakses informasi, mengkaji artikel ilmiah, memetakan, memonitoring, mengarahkan dan mengevaluasi informasi yang terkandung dalam tumpukan dokumen artikel ilmiah tersebut. Maka dari hambatan tersebut, pihak Pusat Penelitian Informatika – LIPI berencana untuk mengelompokkan dokumen menurut karakteristik yang dimiliki setiap artikel ilmiah supaya lebih cepat, tepat waktu dan akurat dalam mendapatkan informasi atau pengetahuan.

Data mining merupakan metode untuk mencari informasi baru yang berguna dari sekumpulan data yang berjumlah besar dan dapat membantu dalam mengambil keputusan. Dalam *data mining* terdapat beberapa metode yaitu estimation, prediction, classification, *clustering* dan association. Dalam tugas

akhir ini akan diterapkan metode *clustering* dengan penggabungan algoritma *hierarchical-clustering* dan algoritma *k-means*. *Clustering* merupakan teknik mengelompokan data dengan melakukan pemisahan data ke dalam sejumlah kelompok menurut karakteristik tertentu, dimana label dari setiap data belum diketahui dan dengan pengelompokan tersebut diharapkan dapat mengetahui kelompok data untuk kemudian diberi label sesuai permasalahan yang dihadapi. Pada dataset yang memiliki banyak obyek dimungkinkan adanya beberapa obyek yang memiliki perbedaan karakteristik dengan yang lainnya, obyek ini disebut *outlier*. Jumlah *outlier* yang sedikit dari banyaknya objek menyebabkan *outlier* sulit untuk terdeteksi. Sedangkan tidak jarang ada informasi penting yang dapat digunakan dari keanomalian tersebut. *Clustering* dapat digunakan untuk menganalisa keberadaa *outlier*.

Oleh sebab itu, Pusat Penelitian Informatika – LIPI memerlukan metode *clustering* untuk dapat menggunakan tumpukan artikel ilmiah tersebut, untuk mendapatkan suatu informasi yang digunakan sebagai acuan dalam mengakses informasi, mengkaji artikel ilmiah, memetakan, memonitoring, mengarahkan dan mengevaluasi informasi.

Dari permasalahan yang ada dan penjelasan yang telah dipaparkan diatas, maka dibutuhkan suatu aplikasi "Penerapan *Analisis outlier* untuk pengelompokkan artikel ilmiah menggunakan metode *clustering*".

I.2 Rumusan Masalah

Berdasarkan latar belakang masalah maka diidentifikasi sebuah masalah yaitu bagaimana cara membangun sebuah sistem menggunakan *data mining* untuk membagi kelompok artikel ilmiah berdasarkan karakteristik yang dimiliki setiap artikel ilmiah untuk pemetaan, monitoring, arahan artikel ilmiah, evaluasi dan menghasilkan informasi dari tumpukan artikel ilmiah tersebut supaya tepat guna bagi masyarakat.

I.3 Maksud dan Tujuan

Berdasarkan permasalahan yang diteliti, maka maksud dari penelitian ini adalah Membangun aplikasi *data mining* yang dapat mengelompokan artikel ilmiah dengan mengimplementasikan metode *clustering*.

Sedangkan tujuan yang akan dicapai dalam penelitian adalah :

1) Memudahkan Pusat Penelitian Informatika – LIPI dalam mengelompokkan artikel ilmiah yang memiliki kesamaan karakteristik khusus oleh setiap artikel ilmiah.
2) Membantu pihak Pusat Penelitian Informatika – LIPI dalam memetakan, monitoring, mengarahkan dan evaluasi.

I.4 Batasan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan sebelumnya, maka dibuat batasan masalah agar penyajian lebih terarah dan mencapai sasaran yang ditentukan.

Adapun batasan masalah yang dibuat adalah sebagai berikut:

1) Data yang akan di analisis adalah artikel ilmiah yang ada di Pusat Penelitian Informatika – LIPI yang akan di transformasi dari *.PDF ke *.TXT.
2) Text/Data yang akan di cluster dari artikel ilmiah adalah bagaian abstrak.
3) Metode yang digunakan adalah *waterfall model.*
4) Sistem yang akan dibangun merupakan aplikasi yang berbasis desktop.
5) Bahasa pemrograman yang akan digunakan adalah Java.
6) Analisis dan pemodelan yang digunakan dalam membangun sistem ini adalah pemodelan berorientasi objek dan dimodelkan dengna UML (*Unified Modeling Language*).

I.5 Metodologi Penelitian

Perancangan dan pembangunan sistem ekstraksi data ke dalam bentuk terstruktur ini menggunakan teknik pengumpulan data sebagai berikut :

I.5.1 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan adalah sebagai berikut :

1. Studi Lapangan

Studi lapangan merupakan teknik pengumpulan data yang dilakukan dengan cara melakukan penelitian ke instansi yang terkait. Studi lapangan ini dilakukan dengan menggunakan 2 cara, yaitu :

a. Observasi

Observasi merupakan kegiatan pengamatan langsung ditempat penelitian untuk mengumpulkan data yang dibutuhkan.

Pada penelitian ini pengamatan yang dilakukan ke Pusat Penelitian Informatika – LIPI.

b. Wawancara

Wawancara merupakan teknik pengumpulan data dengan mengadakan Tanya jawab secara langsung yang berkaitan dengan topik yang diambil.

Pada penelitian ini untuk mendapatkan data – data yang diolah sebagai bahan untuk tugas akhir mewawancarai pihak Pusat Penelitian Informatika – LIPI bagain editor yang bertanggungjawab dalam mengolah jurnal ilmiah.

2. Studi Literatur

Studi literature merupakan teknik pengumpulan literature, jurnal, *paper*, buku – buku, dan bacaan – bacaan yang ada kaitannya dengan penelitian.

I.5.2 Metode pembangunan perangkat lunak

Dalam pembuatan perangkat lunak ini menggunakan *warterfall model* sebagai tahapan pengembangan perangkat lunaknya. Berikut adalah tahapan beserta penjelasan dari *waterfall* model [1] :

1. *Requirement Definition*

   Pada tahapan ini, segala kebutuhan di dalam pembangunan perangkat lunak akan didefinisikan secara lengkap. Kemudian kebutuhan-kebutuhan yang telah terdefinisi tersebut akan dianalisis dan selanjutnya akan ditentukan kebutuhan-kebutuhan apa saja yang harus dipenuhi oleh perangkat lunak yang akan dibangun. Ini merupakan tahapan penting dan harus dikerjakan secara detail agar dapat menghasilkan design yang lengkap.

2. *System and Software Design*

   Pada tahapan ini, setelah seluruh kebutuhan-kebutuhan di dalam pembangunan perangkat lunak telah terkumpul, kemudian perangkat lunak yang akan dibangun akan di desain terlebih dahulu. Desain perangkat lunak akan dibagi berdasarkan bagian-bagiannya dan aktivitas dari bagian-bagian tersebut yang telah dikumpulkan secara lengkap pada tahapan sebelumnya.

3. *Implementation and Unit Testing*

   Pada tahapan ini, hasil dari desain yang telah dibuat akan diterjemahkan ke dalam kode-kode atau *syntax-syntax* pemrograman dengan menggunakan bahasa pemrograman yang telah ditentukan terlebih dahulu dan program yang dibangun akan langsung diuji berdasarkan unit-unitnya atau bagian-bagiannya, supaya tiap-tiap unit atau bagian pada perangkat lunak yang dibangun sesuai dengan tujuan yang ingin dicapai.

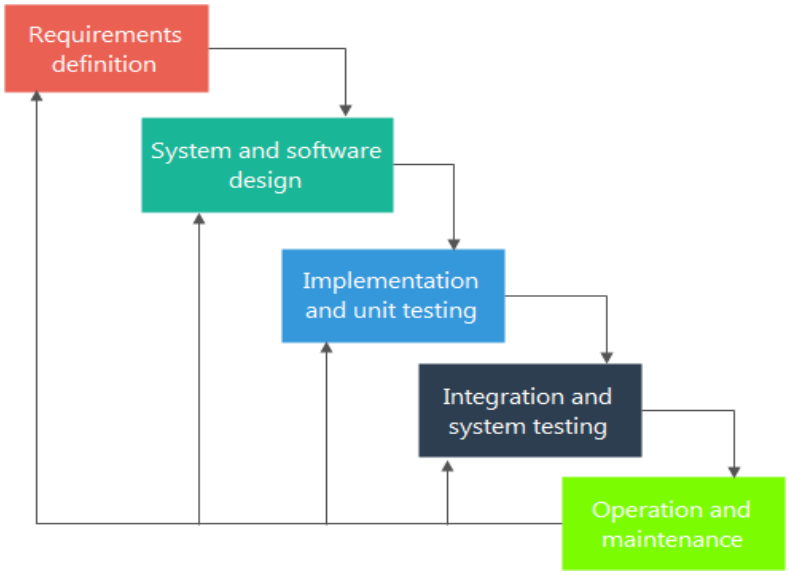
4. *Integration and System Testing*

   Pada tahapan ini seluruh unit-unit program yang terbagi menjadi beberapa bagian-bagian program yang terpisah akan disatukan menjadi satu program atau satu sistem yang utuh, kemudian sistem yang telah

dibuat akan diuji agar dapat berjalan sesuai dengan hasil analisis yang telah dilakukan sebelumnya.

5. *Operation and Maintenance*

Pada tahapan ini, perangkat lunak atau sistem yang telah dibuat akan dilakukan pemeliharaan program, seperti : penyesuaian atau perubahan sistem. Hal ini dilakukan supaya sistem yang telah dibangun dapat beradaptasi dengan situasi yang ada dan juga supaya program atau sistem yang telah dibangun tetap terjaga dari segi fungsionalitasnya. Dari berbagai tahapan-tahapan tersebut, untuk lebih jelasnya bisa dilihat pada Gambar I.1.



Gambar I.1 Waterfall Model

I.6 Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran secara umum mengenai permasalahan dan pemecahannya. Sistematika penulisan tugas akhir ini adalah sebagai berikut :

BAB 1 PENDAHULIAN

Bab ini membahas mengenai latar belakang masalah, identifikasi masalah, maksud dan tujuan, batasan masalah, metode penelitian, pengumpulan data dan metode *data mining* yang digunakan sebagai acuan dalam tahapan penelitian yang akan dilakukan, serta sistematika penulisan untuk menjelaskan pokok – pokok pembahasannya.

BAB 2 LANDASAN TEORI

Pada bab ini akan menjelaskan mengenai objek dari penelitain, dan teori – teori pendukung yang berhubungan dengan pembangunan sistem.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab ini menganalisis masalah dari data hasil penelitian, kemudian dilakukan proses perancangan sistem yang akan dibangun sesuai dengan analisis yang telah dilakukan.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini berisi mengenai implementasi dan analisis dari perancangan sistem yang dilakukan, serta melakukan ujicoba terhadap sistem yang telah dibangun, dan hasil dari uji coba terhadap sistem tersebut.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi mengenai kesimpulan yang diperoleh dari hasil pengujian sistem dan metode yang digunakan, serta saran untuk pengembangan sistem kedepan.