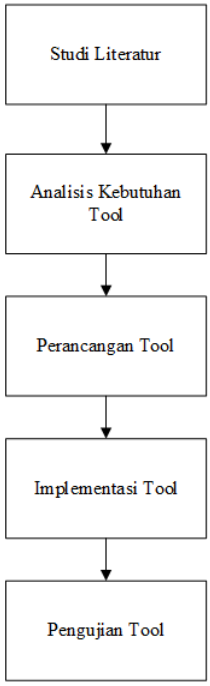
BAB 3
Metode Penelitian

Metode penelitian menjelaskan langkah-langkah yang dilakukan dalam penelitian yang kemudian dijadikan acuan dalam menyelesaikan masalah dalam penelitian, menganalisis hasil penelitian, dan membuat laporan dari hasil penelitian. Alur metode penelitian ditunjukkan pada Gambar 3.1.



Gambar 3.1 Alur Metode Penelitian

3.1 Studi Literatur

Tahapan ini adalah yang pertama kali dilakukan untuk mendapatkan informasi yang akan digunakan sebagai acuan dasar dan penunjang dari tema penelitian yang dilakukan. Informasi-informasi yang dibutuhkan didapatkan dari berbagai sumber, seperti buku, artikel, *paper*, jurnal, makalah, dan laporan penelitian yang didapat secara *online* maupun *offline*. Studi literatur dilakukan untuk mencari informasi yang berkaitan dengan *header email* untuk memahami *value* dari masing-masing *field* yang ada pada *header*.

Pada tahap ini juga dilakukan komparasi *tools* forensik email yang tersedia secara *online* dengan tujuan untuk mengetahui informasi apa saja yang dapat diekstraksi dari *header email* oleh *tools* tersebut. Sehingga fitur dari *tool* yang dibuat dapat melengkapi fitur dari *tools* yang sudah ada sebelumnya.

3.2 Analisis Kebutuhan Tool

Tahapan ini dilakukan untuk menganalisis kebutuhan dari *tool* yang akan dibuat. Ada dua jenis kebutuhan yang dianalisis, yaitu kebutuhan fungsional dan kebutuhan nonfungsional. Kebutuhan fungsional adalah kebutuhan yang berisi proses-proses apa saja yang nantinya dapat dilakukan oleh *tool* yang dibuat. Sedangkan kebutuhan nonfungsional adalah kebutuhan yang menitikberatkan pada properti perilaku yang dimiliki oleh sistem.

3.2.1 Kebutuhan Fungsional

Kebutuhan fungsional dari *tool* yang dibuat adalah *tool* diharapkan dapat membaca masukan berupa satu set *header* dari email. Keluaran yang diharapkan adalah pemetaan informasi dari *header email* tersebut untuk menjawab pertanyaan-pertanyaan:

a. *Who* (siapa pengirim dan penerima email?)
b. *What* (apa subjek dari email? apa file lampiran yang ada pada email?)
c. *When* (kapan email dikirim dan diterima?)
d. *Where* (di mana letak server pengirim email?)
e. *How* (bagaimana proses pengiriman email dari pengirim ke penerima serta *server* apa saja yang dilewati dan protokolnya?)

3.2.2 Kebutuhan Nonfungsional

Kebutuhan nonfungsional terdiri dari perangkat keras dan perangkat lunak yang digunakan dalam membangun *tool* email forensik. Kebutuhan nonfungsionalnya adalah sebagai berikut:

1. *Notebook* Dell Inspiron 7447
2. Sistem Operasi Windows 10
3. Kompiler Netbeans IDE 8.0 untuk pemrograman Java
4. Prosesor Intel Core i7
5. RAM 8GB
6. HDD 1TB

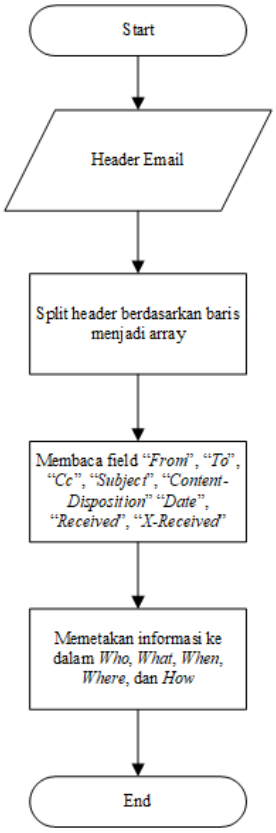
3.3 Perancangan Tool

3.3.1 Perancangan Algoritma Pembacaan Header

Metode yang digunakan dalam membuat *tool* ini adalah dengan membuat algoritma pembacaan *header* dari sebuah email kemudian dilakukan penguraian dengan teknik pencarian berdasarkan kata kunci yang telah ditetapkan. Berikut *field* yang dibutuhkan sebagai kata kunci untuk dapat menjawab pertanyaan-pertanyaan 4W1H:

a. Untuk menjawab *who*, dibutuhkan *field* "*From*", "*To*", dan "*Cc*".
b. Untuk menjawab *what*, dibutuhkan *field* "*Subject*" dan "*Content-Disposition*".
c. Untuk menjawab *when*, dibutuhkan *field* "*Date*" dan "*X-Received*".
d. Untuk menjawab *where*, dibutuhkan *field* "*Received-SPF*".
e. Untuk menjawab *how*, dibutuhkan *field* "*Received*" dan "*X-Received*".

Secara umum, proses ekstraksi informasi dimulai dengan memasukkan *header email*, kemudian *header email* ini akan diuraikan berdasarkan baris menjadi sebuah array. Proses selanjutnya adalah membaca *header email* dari awal dan mencari *field* kata kunci yang dibutuhkan, kemudian isi dari *field* tersebut dipetakan ke *Who*, *What*, *When*, *Where*, dan *How*. Diagram alir proses ekstraksi informasi tersebut ditampilkan pada Gambar 3.2.



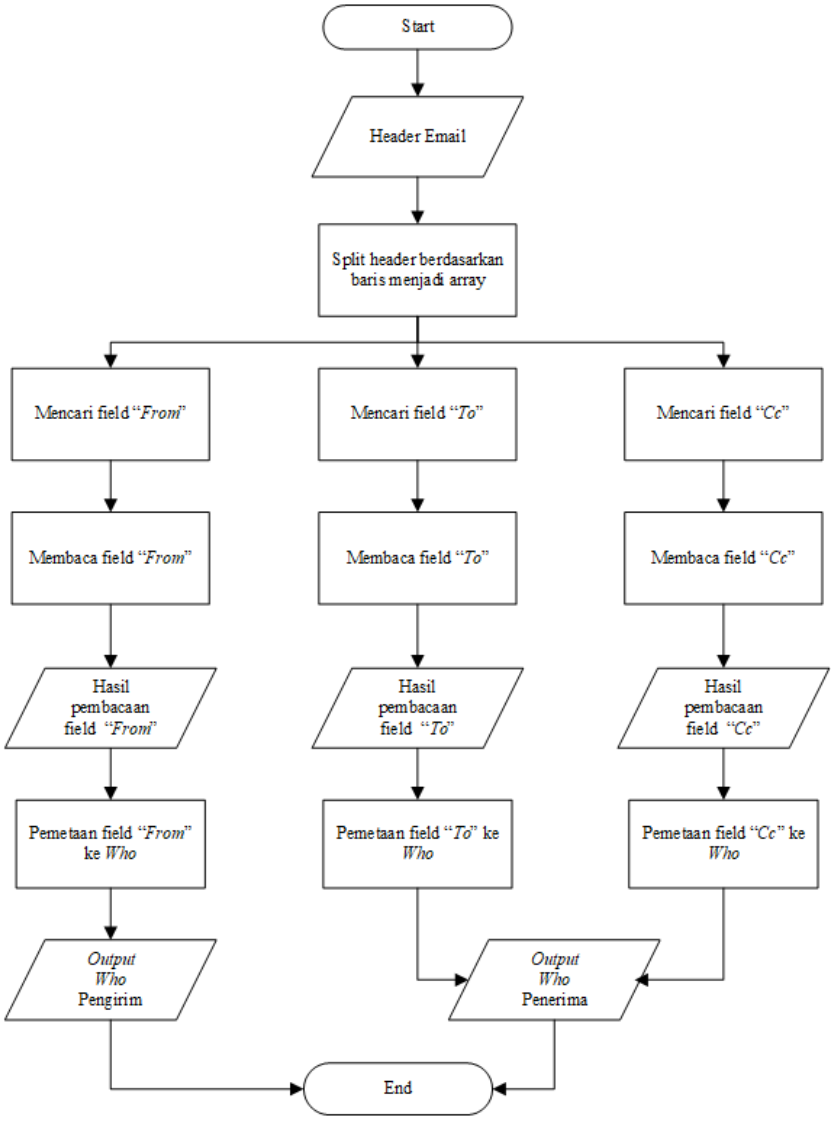
Gambar 3.2 Diagram alir proses ekstraksi informasi

Proses ekstraksi informasi untuk menjawab masing-masing pertanyaan 4W1H, yaitu *Who*, *What*, *When*, *Where*, dan *How*, dijelaskan lebih detail sebagai berikut:

a. Proses ekstraksi informasi untuk menjawab pertanyaan Who

Pertanyaan *Who* dapat menjelaskan siapa saja yang terlibat dalam email tersebut, siapa yang mengirim email, dan siapa yang menerima email. Pada *header email*, informasi mengenai *who* ini didapatkan dari *field From* (pengirim), *To* (penerima), dan *Cc* (penerima). Gambar

3.3 menunjukkan diagram alir proses ekstraksi informasi dari *header email* untuk menjawab pertanyaan *Who*.



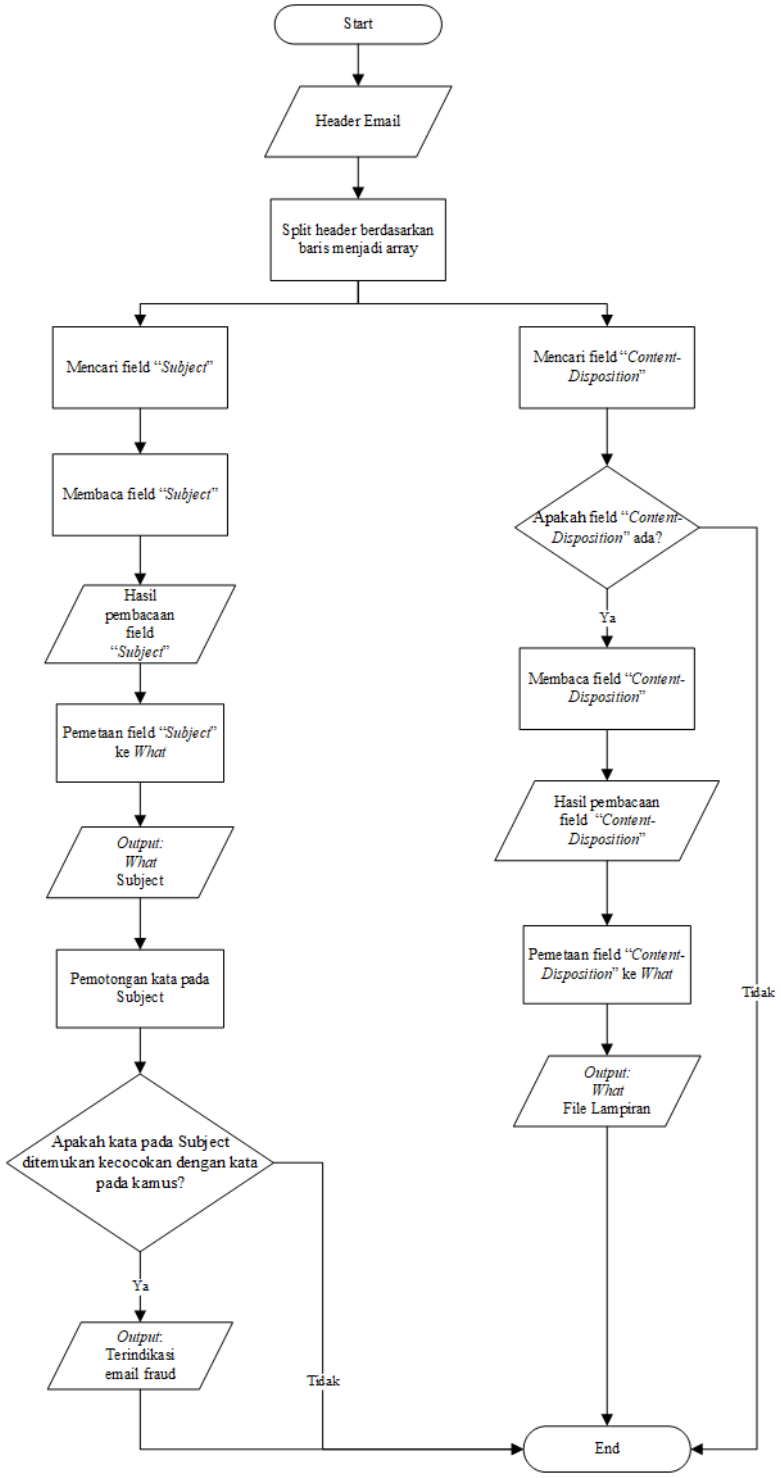
Gambar 3.3 Diagram alir proses ekstraksi informasi untuk menjawab pertanyaan *Who*

Gambar 3.3 menunjukkan bahwa proses ekstraksi informasi untuk menjawab pertanyaan *Who* diawali dengan memasukkan *header email* ke *tool*, kemudian *tool* akan melakukan proses pemecahan *header* berdasarkan baris menjadi sebuah *array*. Selanjutnya *tool* akan mencari *field* yang dibutuhkan, kemudian isi dari *field* tersebut dibaca dan dipetakan ke pertanyaan *Who*.

b. Proses ekstraksi informasi untuk menjawab pertanyaan What

What menunjukkan apa subjek dari email dan file lampiran yang ada pada email. Subjek email dapat menjadi gambaran mengenai isi dari email tersebut. Informasi mengenai subjek didapatkan dari *field Subject* sedangkan informasi mengenai file lampiran didapatkan dari

field Content-Disposition yang terdapat pada *header email*. Proses ektraksi informasi untuk menjawab pertanyaan *What* ditunjukkan oleh Gambar 3.4.



Gambar 3.4 Diagram alir proses ekstraksi informasi untuk menjawab pertanyaan *What*

Berdasarkan Gambar 3.4 di atas, proses ekstraksi informasi untuk menjawab pertanyaan *What* pada prinsipnya hampir sama dengan proses ekstraksi informasi untuk menjawab pertanyaan *Who*. Namun *field* yang dibutuhkan untuk menjawab *What* adalah

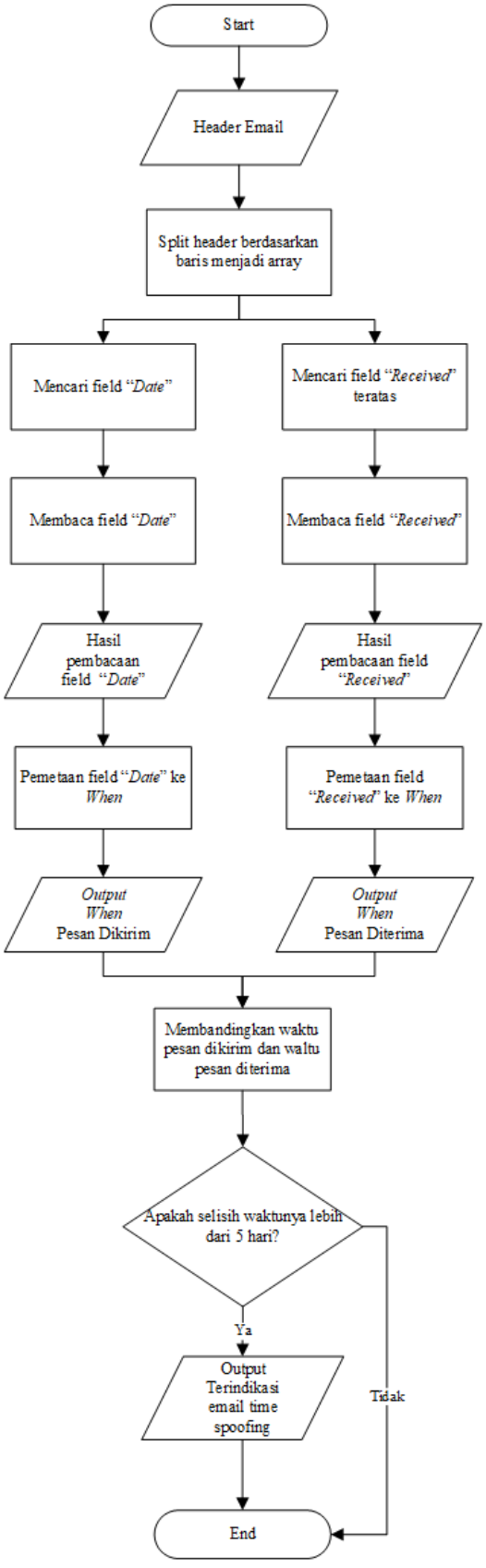
Subject dan *Content-Disposition*. Sehingga *tool* akan mencari kedua *field* tersebut. Setelah *field Subject* ditemukan dan dibaca, kemudian dipetakan ke pertanyaan *What*. Saat isi dari *field Subject* telah berhasil dipetakan, kemudian dilakukan proses pemotongan kata untuk dibandingkan dengan kata pada kamus. Proses ini dilakukan untuk mendeteksi apakah terdapat indikasi bahwa email tersebut merupakan *email fraud* atau bukan.

Field selanjutnya yang dicari adalah *Content-Disposition*, yang mana *field* ini tidak selalu ada pada *header email*. *Field* ini akan muncul ketika dalam email terdapat *file* yang dilampirkan bersama isi email. Ketika ditemukan *field Content-Disposition*, informasi *filename* yang ada pada *field* ini dibaca dan dipetakan ke *What*. Namun ketika *field* ini tidak ditemukan, maka informasinya tidak dipetakan.

c. Proses ekstraksi informasi untuk menjawab pertanyaan When

Pertanyaan *When* menunjukkan waktu email dikirim oleh pengirim dan waktu email diterima oleh penerima. Informasi mengenai waktu pengiriman dan penerimaan email didapatkan dari *field Date* (waktu pengiriman) dan *Received* (waktu penerimaan) yang terakhir atau teratas dari *header email*. Gambar 3.5 menampilkan proses ekstraksi informasi dari *field Date* dan *Received* untuk menjawab pertanyaan *When*. Proses ekstraksinya hampir sama dengan proses ekstraksi untuk menjawab *Who* dan *What*. Perbedaan terdapat pada *field* yang diambil informasinya, yaitu *field Date* dan *Received* yang diperlukan. *Tool* akan mencari *field Date* dan *Received* yang teratas. Setelah informasi dari kedua *field* ini dibaca, jika informasi waktu yang tercantum bukan dalam GMT +7 atau WIB (Waktu Indonesia Barat), akan dilakukan konversi terlebih dahulu ke dalam WIB, kemudian baru dipetakan ke *When*.

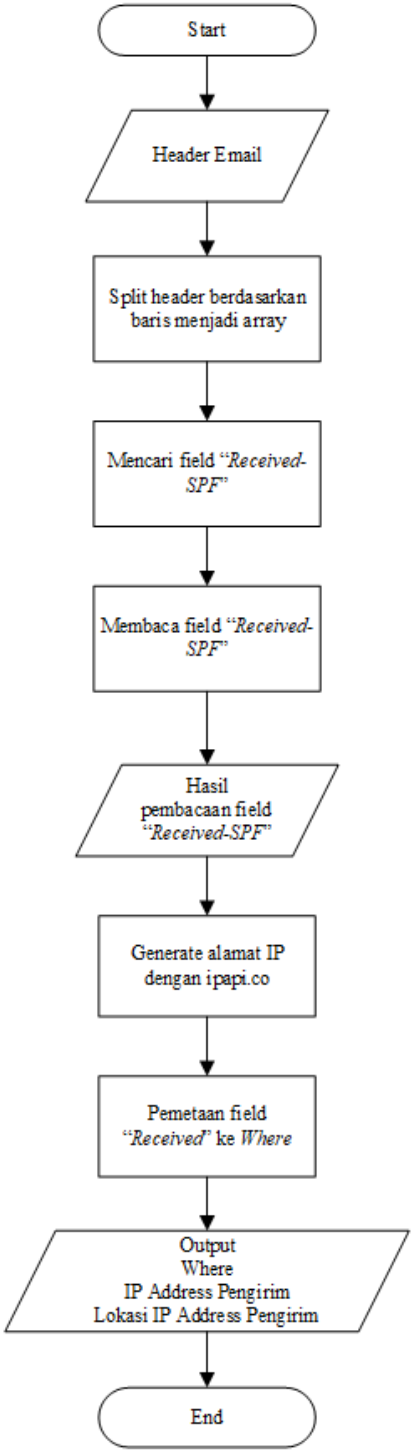
Proses selanjutnya yaitu dilakukan pembandingan antara waktu email dikirim oleh *server* pengirim dengan waktu email diterima oleh *server* penerima. Hal ini dilakukan untuk mendeteksi *email time spoofing* terhadap *header email* yang diujikan di *tool*. Proses-proses tersebut ditunjukkan oleh Gambar 3.5.



Gambar 3.5 Diagram alir proses ekstraksi informasi untuk menjawab pertanyaan *When*

d. Proses ekstraksi informasi untuk menjawab pertanyaan Where

Pertanyaan *Where* menunjukkan alamat IP dari *server* pengirim email beserta lokasinya. Informasi alamat IP ini didapatkan dari *field Received-SPF* pada *header email*. Proses ekstraksi informasi dari *field Received-SPF* untuk menjawab pertanyaan *Where* ditunjukkan pada Gambar 3.6.

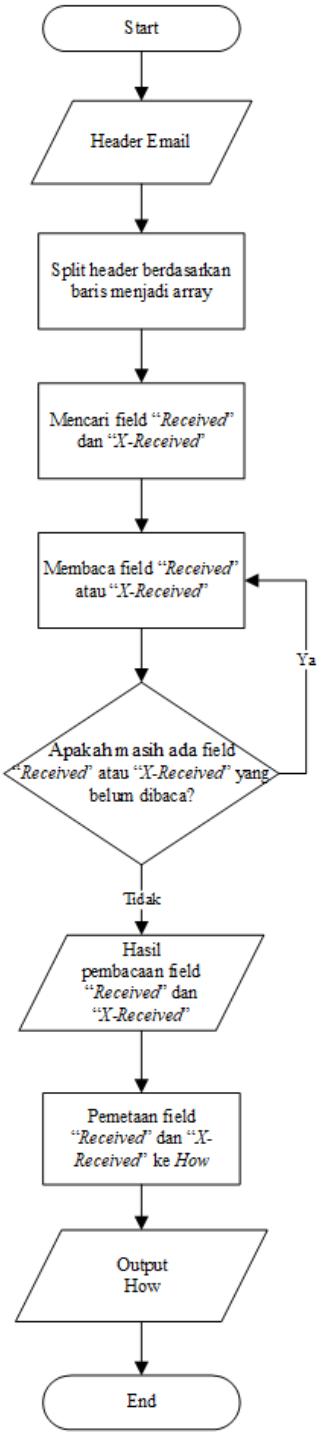


Gambar 3.6 Diagram alir proses ekstraksi informasi untuk menjawab pertanyaan *Where*

Gambar 3.6 menunjukkan diagram alir proses ekstraksi informasi untuk menjawab pertanyaan *Where*. *Field* yang diekstraksi adalah *Received-SPF* untuk mendapatkan informasi alamat IP dari *server* pengirim email. Proses ekstraksi dimulai setelah *header email* dimasukkan ke *tool,* kemudian *header email* tersebut diurai menjadi suatu array berdasarkan baris. Selanjutnya *tool* akan mencari *field* yang dibutuhkan, yaitu *Received-SPF* kemudian isi dari *field* tersebut dibaca. Setelah didapatkan alamat IP-nya kemudian alamat IP ini di-*generate* melalui ipapi.co untuk mengetahui lokasi dari alamat IP ini. Proses terakhir, alamat IP dan lokasi dari alamat IP ini dipetakan ke *Where*.

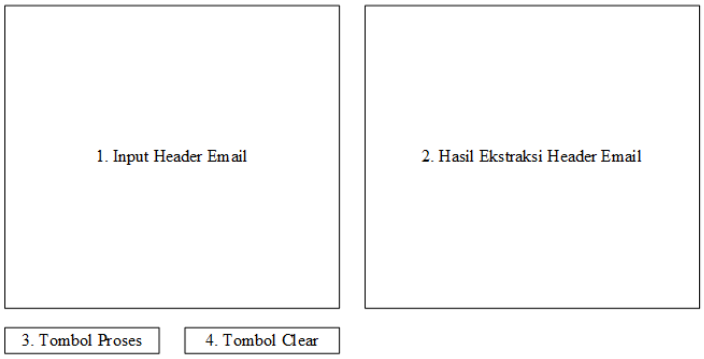
e. Proses ekstraksi informasi untuk menjawab pertanyaan How

Pertanyaan *How* menunjukkan proses perjalanan email dari pengirim ke penerima, *server* yang dilewati oleh email dan protokol yang digunakan selama di perjalanan. Informasi untuk menjawab pertanyaan *How* diambil dari *field Received* dan *X-Received*. Proses ekstraksi informasinya sama dengan proses ekstraksi informasi untuk menjawab pertanyaan *Who, What, When,* dan *Where*. *Tool* akan mulai bekerja ketika ada masukan berupa *header email*. Kemudian *tool* memecah *header email* tadi berdasarkan baris menjadi sebuah array. Setelah itu *tool* akan mencari semua *field Received* dan *X-Received* dan membaca isi dari masing-masing *field* tersebut, yaitu *from, by* dan protokolnya. *Field Received* dan *X-Received* dibaca dari paling bawah, setelah selesai membaca satu *field*, kemudian dicek lagi apakah masih ada *field Received* atau *X-Received* yang belum terbaca, jika masih ada, maka *tool* akan kembali mencari *field Received* atau *X-Received* yang belum dibaca. Begitu seterusnya sampai semua *field Received* dan *X-Received* terbaca, kemudian dipetakan ke *How*. Diagram alir proses ekstraksi informasi untuk menjawab pertanyaan *How* ditunjukkan pada Gambar 3.7.



Gambar 3.7 Diagram alir proses ekstraksi informasi untuk menjawab pertanyaan *How*

3.3.2 Tampilan antarmuka *tool* Mail Header Extractor



Gambar 3.8 Tampilan antarmuka *tool* Mail Header Extractor

Gambar 3.8 merupakan rancangan tampilan antarmuka *tool* Mail Header Extractor yang dibuat pada penelitian ini. Berikut adalah penjelasan dari masing-masing bagian:

1. Input Header Email, adalah tempat untuk memasukkan *header email* yang ingin diekstrak informasinya.
2. Hasil Ekstraksi Header Email, adalah tempat untuk menampilkan informasi hasil pemetaan dari *header email* yang telah dimasukkan sebelumnya.
3. Tombol Proses, tombol untuk memulai proses ekstraksi informasi dari *header email*.
4. Tombol Clear, tombol untuk menghapus kolom header email dan kolom hasil ekstraksi untuk memasukkan *header email* lain.

3.3.3 Fitur Mail Header Extractor

Berikut ini adalah beberapa fitur yang ada pada *tool* Mail Header Extractor, yaitu:

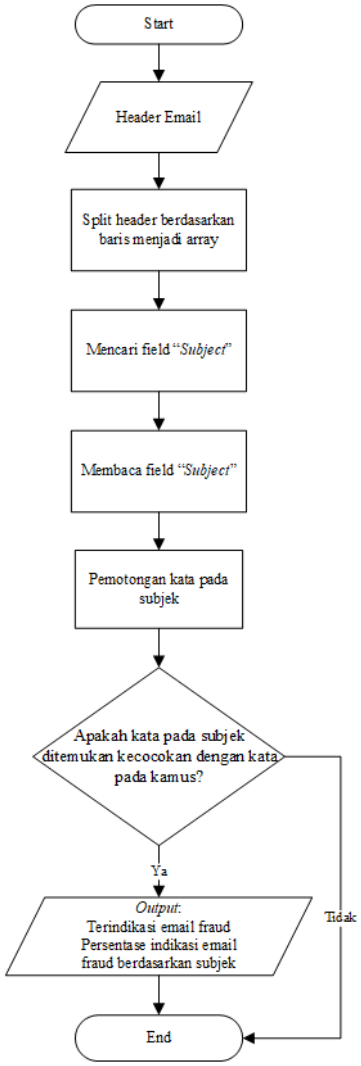
a. Deteksi *email fraud*

   Fitur ini ditambahkan untuk memberikan indikasi jika *header email* yang diujikan diduga sebagai *email fraud*. Proses deteksi *email fraud* dilakukan dengan dua cara, yaitu dengan mengecek subjek email dan isi email.

   1. Deteksi *email fraud* berdasarkan subjek email

      Prosesnya dilakukan dengan membandingkan kata pada subjek email yang diujikan dengan kata pada kamus. Kamus ini disusun berdasarkan dataset Nigerian Fraud Letters. Dataset tersebut memuat 4291 email yang dikumpulkan dari tahun 1998 hingga 2007. Subjek email dari 4291 email tersebut dikumpulkan kemudian dibuat pemeringkatan berdasarkan kata yang paling sering muncul pada subjek tersebut. Jika ada kata pada subjek email yang cocok dengan kata pada kamus, maka email tersebut

diindikasikan sebagai *email fraud*. Diagram alir proses deteksi *email fraud* berdasarkan subjek email ditunjukkan pada Gambar 3.9.



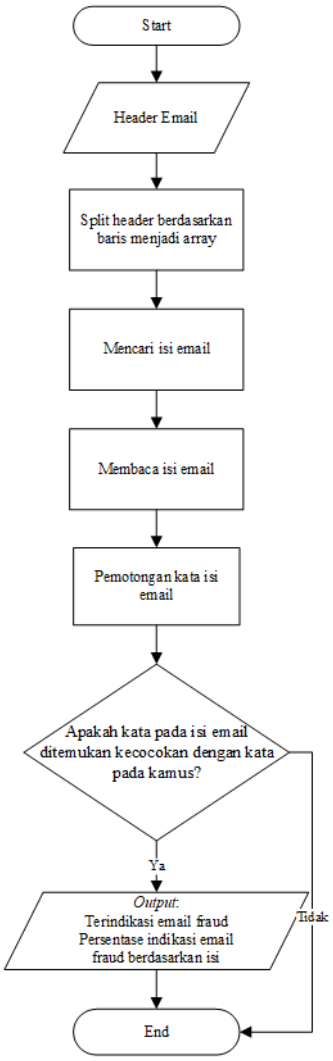
Gambar 3.9 Diagram alir proses deteksi *email fraud* berdasarkan subjek email

Gambar 3.9 menunjukkan diagram alir proses deteksi *email fraud* berdasarkan subjek email. Saat *field Subject* telah berhasil dibaca, kemudian dilakukan pemotongan kata pada subjek tersebut dan setiap katanya dicocokkan dengan kata yang ada pada kamus. Jika ada kata yang cocok, maka diindikasikan sebagai *email fraud* dengan angka persentase yang menunjukkan perbandingan antara jumlah kata pada subjek yang cocok dengan kata pada kamus.

2. Deteksi *email fraud* berdasarkan isi email

   Proses deteksinya hampir sama dengan proses deteksi berdasarkan subjek email. Perbedaan terletak pada proses penyusunan data untuk kamusnya. Untuk mendapatkan kamus yang menjadi pembanding isi email, diambil 25 sampel email dari dataset

Nigerian Fraud Letters yang kemudian dibuat pemeringkatan berdasarkan kata yang paling sering muncul dari isi 25 email tersebut. Jika ada kata pada isi email yang cocok dengan kata pada kamus ini, maka email akan diindikasikan sebagai *email fraud*. Gambar 3.10 menggambarkan diagram alir proses deteksi *email fraud* berdasarkan isi email.



Gambar 3.10 Diagram alir proses deteksi *email fraud* berdasarkan isi email

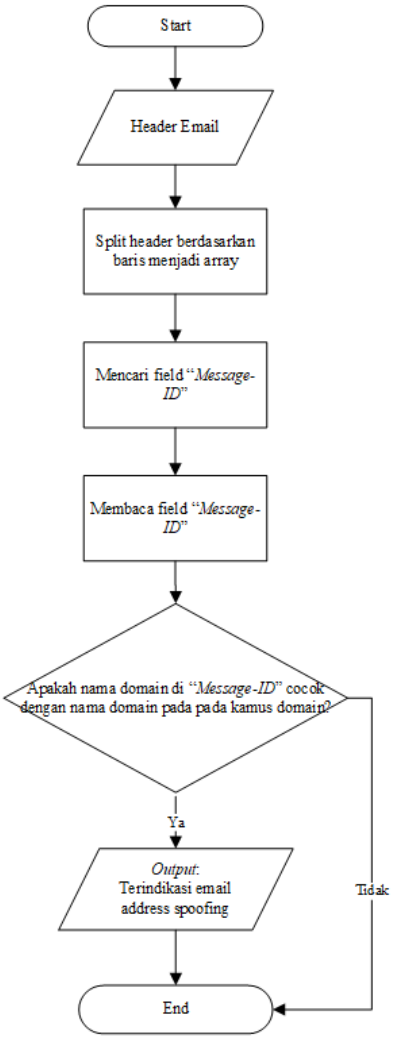
Berdasarkan Gambar 3.10 di atas, proses deteksi *email fraud* berdasarkan isi email dilakukan dengan cara membaca isi email kemudian memotong setiap kata pada isi email. Selanjutnya setiap kata tadi dicocokkan dengan kata yang ada pada kamus yang telah dibuat sebelumnya. Jika terdapat kata pada isi email yang cocok dengan kata yang ada pada kamus, email yang diujikan diindikasikan sebagai *email fraud*. Ditampilkan pula persentasenya yang angkanya didapatkan dari perbandingan antara jumlah kata pada isi email yang cocok dengan kata pada kamus.

b. Deteksi *email spoofing*

Fitur ini dibuat untuk mendeteksi adanya indikasi pemalsuan yang dilakukan dengan mengubah isi dari *field* yang ada pada *header email*. Terdapat dua kemungkinan manipulasi yang dilakukan, yaitu *address spoofing* (memalsukan alamat email pengirim dengan mengubah *field From* pada *header email*) dan *time spoofing* (memalsukan tanggal pengiriman dengan mengubah *field Date* yang ada pada *header email*).

1. Deteksi *email address spoofing*

   Proses deteksinya dilakukan dengan membandingkan nama domain yang ada pada *field Message-ID* pada *header email* yang diujikan dengan nama domain yang ada pada kamus domain. Data nama domain pada kamus domain ini didapatkan dari hasil simulasi pengiriman *email spoofing* dari berbagai situs *fake mailer* yang tersedia di internet. Jika nama domain pada *field Message-ID* cocok dengan nama domain yang ada pada kamus domain, maka email diindikasikan sebagai *email address spoofing*. Diagram alir proses deteksi *email address spoofing* digambarkan pada Gambar 3.11.

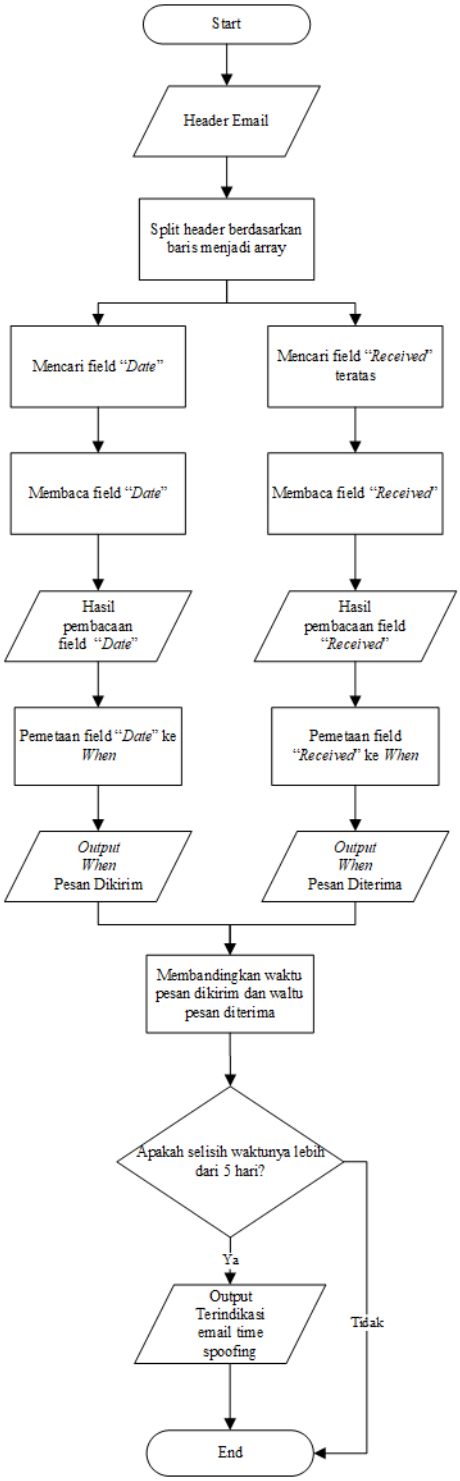


Gambar 3.11 Diagram alir proses deteksi *email address spoofing*

Gambar 3.11 menunjukkan diagram alir proses deteksi *email address spoofing* yang dilakukan oleh *tool* Mail Header Extractor. *Tool* akan mencari *field Message-ID* dari *header email* yang diujikan dan membaca nama domain yang tercantum, kemudian membandingkannya dengan nama domain yang ada pada kamus domain. Jika nama domain pada *field Message-ID* ditemukan kesamaan dengan nama domain pada kamus domain, maka email akan diindikasikan sebagai *email address spoofing*.

2. Deteksi *email time spoofing*

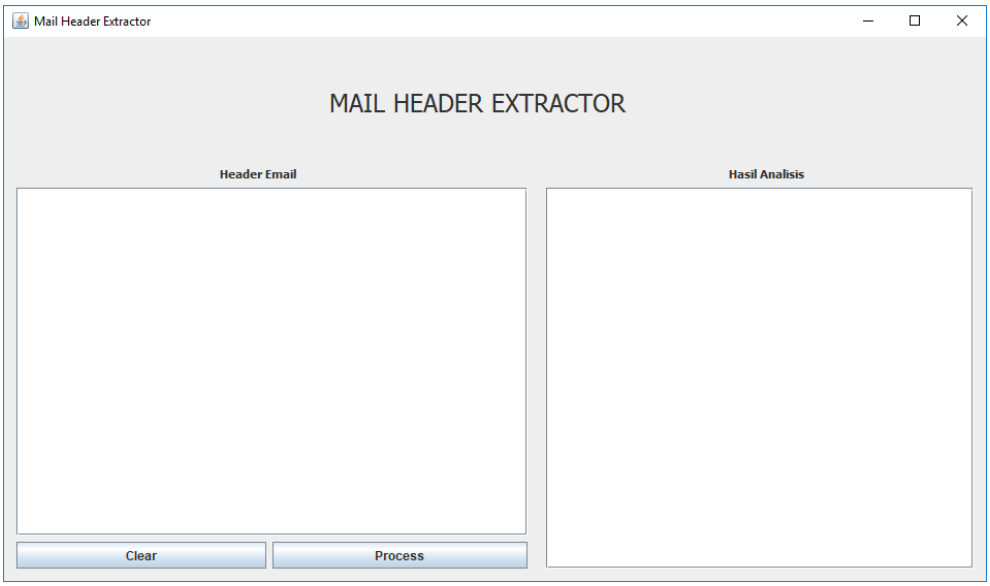
   Proses pendeteksiannya dilakukan dengan membandingkan waktu pengiriman email (didapatkan dari *field Date*) dan waktu penerimaan email (didapatkan dari *field Received* atau *X-Received* yang teratas). Sebelum dilakukan proses pembandingan, waktu yang tercatat pada *field* tersebut dikonversi ke dalam WIB (Waktu Indonesia Barat) terlebih dahulu. Jika selisih antara waktu pengiriman email dengan waktu penerimaan email lebih dari 5 hari, maka email akan diindikasikan sebagai *email time spoofing* (Klensin, 2008). Gambar 3.12 menampilkan diagram alir proses deteksi *email time spoofing*.



Gambar 3.12 Diagram alir proses deteksi *email time spoofing*

3.4 Implementasi Tool

Implementasi adalah proses penerapan rancangan yang telah dibuat sebelumnya. Implementasi *tool* dilakukan dengan menggunakan bahasa pemrograman Java dengan *compiler* Netbeans 8.0. Gambar 3.13 menunjukkan tampilan dari *tool* Mail Header Extractor yang dibuat.



Gambar 3.13 Tampilan *tool* Mail Header Extractor

Kolom Header Email adalah tempat untuk memasukkan *header email* yang diujikan. Tombol *Process* di sebelah kanan bawah, di bawah kolom Header Email adalah tombol untuk memulai proses ekstraksi informasi dari *header email* yang sudah dimasukkan. Setelah tombol *Process* ditekan, *tool* akan mulai melakukan pembacaan *header email*, mencari *field* kata kunci, membaca isinya, kemudian memetakan informasinya ke dalam 4W1H dan menampilkan hasilnya pada kolom Hasil Analisis. Tombol *Clear* berfungsi untuk menghapus isi dari kolom Header Email dan kolom Hasil Analisis.

3.5 Pengujian Tool

Pengujian *tool* dilakukan dengan menggunakan beberapa skenario pengujian. Pengujian ini bertujuan untuk menilai apakah *tool* telah berhasil mengekstraksi setiap masukan *header email* dan memetakan informasinya ke dalam konsep 4W1H. Sejumlah skenario yang ditetapkan adalah sebagai berikut:

a. Email normal.
b. Email yang subjeknya berupa *fraud*.
c. Email yang telah dipalsukan alamat pengirimnya.
d. Email yang telah dipalsukan tanggal pengirimannya.
e. Email yang telah dipalsukan alamat pengirim dan tanggal pengirimannya.
f. Email yang subjeknya berupa *fraud* serta telah dipalsukan alamat pengirim dan tanggal pengirimannya.