

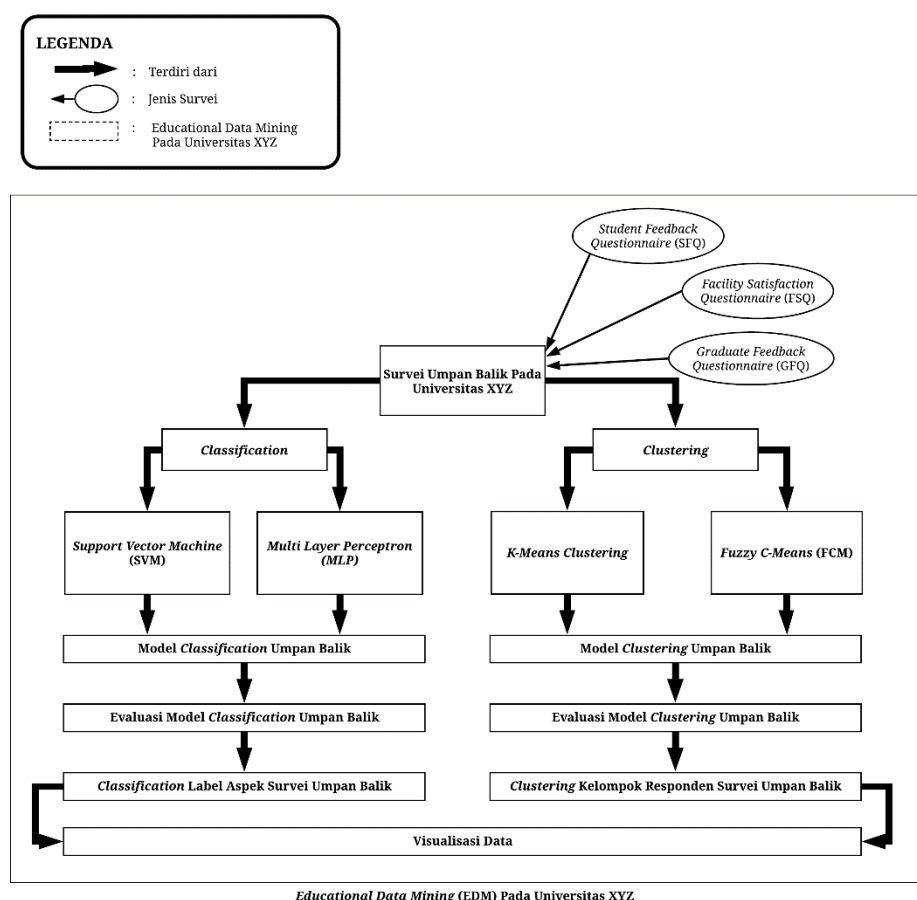
## BAB III

### METODOLOGI PENELITIAN

Bab ini membahas metode yang digunakan untuk menjawab pertanyaan penelitian yang terdiri dari metodologi penelitian berisi kerangka teoritis yang menjadi dasar penelitian, tahapan penelitian, metode pengumpulan data, metode analisis data dan perancangan alur pengembangan model serta penggunaannya.

#### 3.1 Metodologi Penelitian

Pada bagian ini dijabarkan *theoretical framework* yang berisi tahapan penelitian berdasarkan penggunaan teori-teori. Tahapan tersebut dapat dilihat pada Gambar 3.1.



Gambar 3.1 *Theoretical Framework* Pengolahan Survei Umpan Balik

Gambar 3.1 menunjukkan *theoretical framework* pada penelitian ini. Bidang *data mining* pada penelitian ini adalah *Educational Data Mining* (EDM) yaitu salah satu bidang *data mining* untuk mendapatkan pengetahuan dalam bidang pendidikan [8]. EDM yang dikembangkan adalah EDM dalam lingkup Universitas XYZ.

Dalam dokumen Peraturan Permendikbud Nomor 5 Tahun 2020, pemangku kepentingan universitas berperan penting di dalam universitas. Pemangku kepentingan universitas antara lain mahasiswa. Oleh karena itu diambil data survei umpan balik kepada mahasiswa untuk mengolah data yang berisi pendapat mereka. Survei umpan balik pada universitas yang akan dilakukan pada penelitian ini yaitu *Student Feedback Questionnaire* (SFQ), *Facility Satisfaction Questionnaire* (FSQ), dan *Graduate Feedback Questionnaire* (GFQ).

Dalam penelitian ini pengolahan data survei umpan balik pada universitas dilakukan dengan menggunakan teknik *classification* dan *clustering* berdasarkan teknik yang digunakan pada penelitian sebelumnya. Teknik *classification* digunakan untuk mengolah data survei umpan balik yang berbentuk *Open Ended Question* (OEQ), sedangkan teknik *clustering* digunakan untuk mengolah data survei umpan balik dalam bentuk *Close Ended Question* (CEQ). Algoritma yang digunakan dalam mengolah survei umpan balik pada teknik *classification* adalah *Support Vector Machine* (SVM) dan *Multi Layer Perceptron* (MLP). Pengolahan data survei umpan balik pada teknik *clustering* adalah *K-Means Clustering* dan *Fuzzy C-Means Clustering* (FCM). Algoritma-algoritma tersebut dipilih untuk digunakan dalam penelitian ini berdasarkan algoritma yang digunakan pada penelitian-penelitian sebelumnya.

Penelitian ini melakukan teknik *classification* untuk memprediksi label aspek pada OEQ dalam jenis survei SFQ, FSQ, dan GFQ. Model yang dihasilkan dengan *classification* dievaluasi dengan menghitung *precision*, *accuration*, *recall*, dan *F1-Score* dari setiap model. Selain melakukan teknik *classification*, pada penelitian ini dilakukan teknik *clustering* untuk mendapatkan *cluster* atau kelompok responden pada CEQ dalam jenis survei SFQ, FSQ, dan GFQ. Model yang dihasilkan dengan *clustering* dievaluasi dengan menghitung *silhouette coefficient*. Evaluasi dilakukan agar mendapatkan model *classification* dan model *clustering* yang terbaik dalam

penelitian ini. Pada penelitian ini juga dilakukan visualisasi data agar pengetahuan dari data yang dihasilkan dapat dikomunikasikan dengan baik kepada pengguna.

### 3.2 Tahapan Penelitian

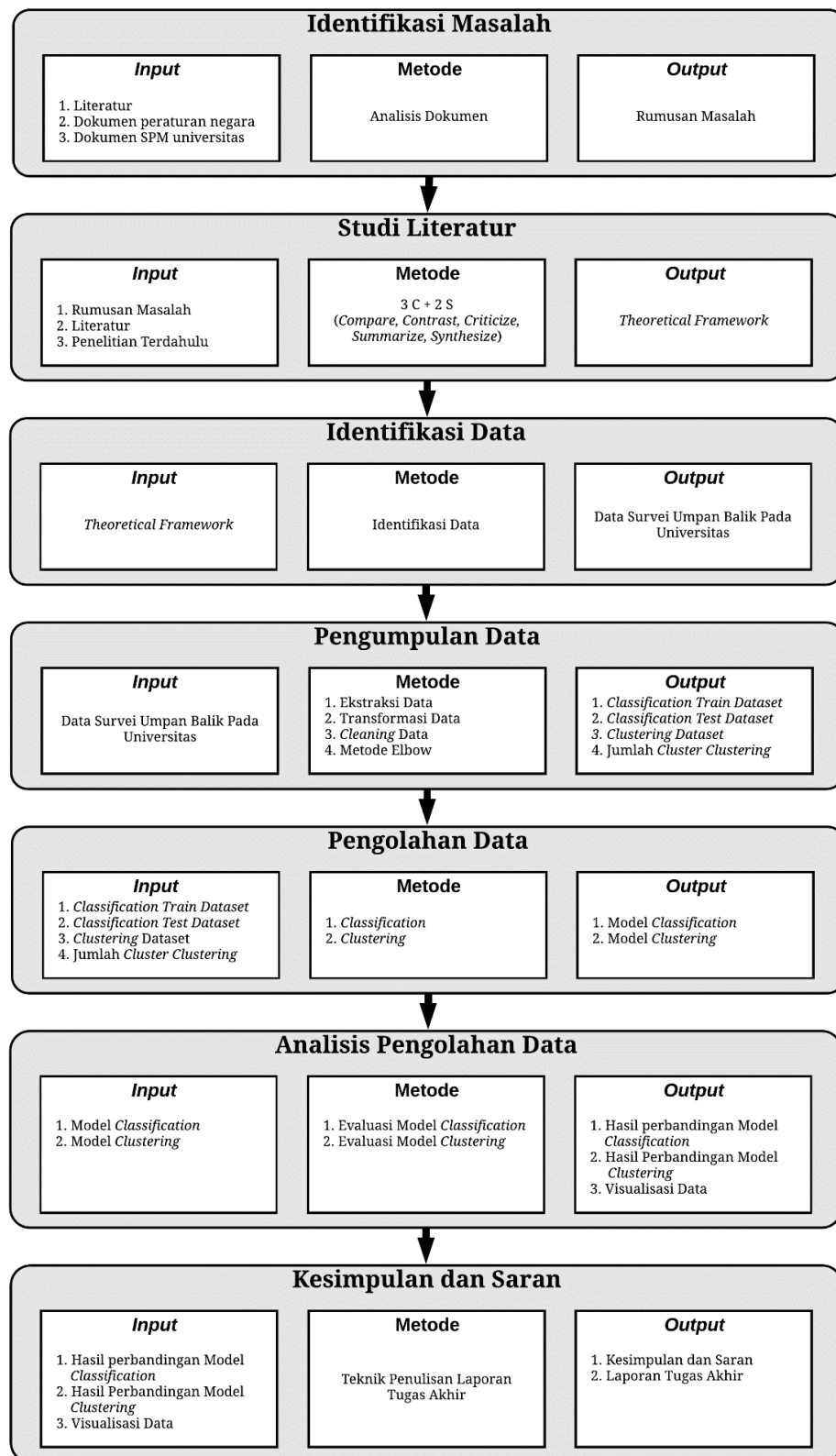
Tahapan penelitian berisi tahap acuan dalam melakukan penelitian. Tahapan penelitian berisi alur penelitian dengan menjabarkan *input*, metode, dan *output* pada setiap tahap. Gambar 3.2 menunjukkan tahapan alur penelitian yang digunakan pada penelitian ini. Tahapan penelitian dimulai dari adanya kebutuhan akan pengontrolan kualitas pada pengolahan data survei umpan balik di universitas. Selain itu cara pengolahan data yang masih manual turut menjadi masalah karena dapat meningkatkan potensi ketidakakuratan analisis. Berdasarkan kepentingan tersebut dihasilkan rumusan masalah yaitu: “Bagaimana pemanfaatan *Educational Data Mining* (EDM) dalam pengolahan data survei umpan balik mahasiswa yang memiliki bentuk *Open Ended Question* (OEQ) digabungkan dengan bentuk *Close Ended Question* (CEQ)?”. Tahapan penelitian dirancang dilakukan untuk menjawab rumusan masalah tersebut.

Tahapan selanjutnya adalah pemilihan teknik yang dapat digunakan untuk melakukan teknik *classification* dan *clustering* dengan mengacu pada studi literatur pada penelitian terdahulu. Seluruh tahapan tersebut menghasilkan sebuah *Theoretical Framework* penelitian. Setelah tahap tersebut, tahap penelitian dilanjutkan dengan melakukan identifikasi terhadap data survei umpan balik yang digunakan pada teknik *classification* serta teknik *clustering*. Hasil dari tahap tersebut adalah data untuk melakukan teknik *classification* dan *clustering*.

Tahapan berikutnya adalah pengelolaan data dengan melakukan *data cleaning*, *data integration*, *data selection*, *data transformation*, dan *data mining* untuk mendapatkan *train dataset* dan *test dataset* pada teknik *classification* serta *dataset* untuk teknik *clustering*. Model teknik *classification* didapatkan dengan menggunakan *train dataset* dan *test dataset* terhadap data OEQ. Model teknik *clustering* didapatkan dengan menggunakan *dataset* berbentuk skor jawaban setiap responden survei CEQ dan jumlah *cluster* yang telah ditentukan menggunakan *Elbow method*. *Train dataset* digunakan untuk memperoleh model teknik

*classification* terhadap data survei umpan balik dalam bentuk OEQ pada jenis survei SFQ, FSQ, dan GFQ. *Testing dataset* pada *classification* digunakan untuk melakukan pengujian hasil dari model teknik *classification*. *Dataset* dengan bentuk skor jawaban setiap responden survei CEQ serta penentuan jumlah *cluster* digunakan untuk memperoleh model teknik *clustering* terhadap survei umpan balik dalam bentuk CEQ pada jenis survei SFQ, FSQ, dan GFQ. Hasil dari *clustering* tersebut yang akan digunakan untuk evaluasi model *clustering*.

Tahap selanjutnya adalah melakukan uji *precision*, *accuration*, *recall*, dan *F1-Score* untuk model *classification* dan *silhouette coefficient* untuk model *clustering*. Evaluasi menghasilkan model terbaik pada *classification* label aspek survei SFQ, FSQ, dan GFQ untuk menghasilkan label aspek yang lebih akurat. Hasil evaluasi pada *clustering* digunakan untuk memilih model terbaik pada *clustering*, yaitu algoritma yang lebih tepat, jumlah *cluster* yang lebih optimal dan hasil analisis yang lebih tepat. Tujuan dari hasil evaluasi adalah untuk menetapkan model yang terbaik sehingga menghasilkan label aspek *classification* dan hasil pengelompokan responden yang terbaik. Setelah itu dilakukan visualisasi data untuk mempermudah eksplorasi hasil. Kesimpulan dan saran penelitian diambil setelah data-data diolah dan dianalisis terlebih dahulu.



Gambar 3.2 Tahapan penelitian

### 3.3 Metode Pengumpulan Data

Survei umpan balik pada universitas yang digunakan dalam penelitian ini adalah survei umpan balik kepada mahasiswa. Berikut ini adalah penjelasan singkat mengenai jenis survei yang digunakan dalam penelitian ini.

*Student Feedback Questionnaire* (SFQ) merupakan survei umpan balik yang dilakukan secara berkala kepada mahasiswa Universitas XYZ untuk mengukur dan memperoleh umpan balik terhadap dosen. Survei SFQ dilakukan pada setiap semester yaitu sebanyak tiga kali dalam setahun. Aspek-aspek yang menjadi indikator pada survei SFQ yaitu:

- 1) Penyedia Informasi (*Information Provider*)
- 2) Panutan (*Role Model*)
- 3) Fasilitator (*Facilitator*)
- 4) Penilai (*Assesor*)
- 5) Bahasa Inggris (*English*)
- 6) Pembelajaran Mahasiswa (*Student Learning*)

*Facility Satisfaction Questionnaire* (FSQ) merupakan survei umpan balik yang dilakukan secara berkala kepada mahasiswa untuk mengukur umpan balik mahasiswa terhadap fasilitas dan jasa yang disediakan Universitas XYZ. Survei umpan balik FSQ dilakukan setiap tahun sebanyak satu kali. Aspek-aspek yang menjadi indikator pada survei FSQ yaitu:

- 1) Fasilitas dan Perlengkapan Olahraga
- 2) Fasilitas dan Perlengkapan Lab
- 3) Fasilitas Ruang UKM
- 4) Fasilitas *Student Lounge*
- 5) Ruangan dan Perlengkapan Kelas
- 6) Fasilitas Layanan Konseling
- 7) Fasilitas dan Layanan Kesehatan
- 8) Fasilitas dan Layanan Kantin
- 9) Fasilitas dan Layanan Parkir



- 10) Fasilitas dan Layanan IT
- 11) Kebersihan Lingkungan Kampus
- 12) Layanan Pembimbing Akademik
- 13) Layanan Pengembangan Kerohanian
- 14) Layanan *Student Life*
- 15) Layanan Admisi dan *Marketing*
- 16) Layanan International Office
- 17) Layanan Keuangan
- 18) Layanan Keamanan
- 19) Layanan ADAK
- 20) Layanan Tata Usaha Fakultas/Jurusan

*Graduate Feedback Questionnaire* (GFQ) merupakan survei umpan balik yang dilakukan secara berkala kepada mahasiswa yang baru lulus untuk mengukur pengalaman mahasiswa saat melakukan perkuliahan di Universitas XYZ. Survei umpan balik GFQ dilakukan pada semester ganjil dan semester genap yaitu sebanyak dua periode kelulusan dalam setahun. Aspek-aspek yang menjadi indikator dalam GFQ adalah sebagai berikut:

- 1) Proses Pembelajaran
- 2) Pembimbing Tugas Akhir
- 3) Pembimbing Akademik
- 4) Komputer Lab, Studio, *Workshop*
- 5) Kelas dan Pendukung Kelas
- 6) Fasilitas IT
- 7) Layanan Perpustakaan
- 8) ADAK & Keuangan
- 9) Organisasi Kemahasiswaan
- 10) Bimbingan dan Konseling
- 11) Minat dan Bakat
- 12) Pengembangan *Soft Skill*
- 13) Layanan Beasiswa

#### 14) Layanan Kesehatan

Data yang digunakan pada penelitian ini adalah data survei umpan balik pada universitas yang didapatkan secara *online* dalam bentuk *file Microsoft Excel*. Lampiran A merupakan contoh survei umpan balik SFQ, Lampiran B merupakan contoh survei umpan balik FSQ, dan Lampiran C merupakan contoh survei umpan balik GFQ. Data aspek yang akan dijadikan label aspek atau kategori aspek akan diatur kembali untuk melakukan penelitian ini. Tabel 3.1 menunjukkan rekapitulasi aspek dan pertanyaan survei umpan balik pada universitas.

Tabel 3.1 Rekapitulasi Aspek dan Pertanyaan Survei Umpan Balik

Jenis Survei	Jumlah Pertanyaan <i>Open Ended Question</i>	Jumlah Pertanyaan <i>Close Ended Question</i>	Jumlah Aspek
SFQ	3	19	6
FSQ	1	40	20
GFQ	1	28	14

### 3.4 Metode Analisis Data

Penelitian ini menggunakan teknik *classification* untuk mendapatkan model *classification* label aspek survei umpan balik pada universitas dalam bentuk *Open Ended Question* (OEQ) dengan tipe survei SFQ, FSQ dan GFQ. Penelitian ini juga menggunakan teknik *clustering* untuk mendapatkan model *clustering* dalam pengelompokan responden survei umpan balik pada universitas dalam bentuk *Close Ended Question* (CEQ) dalam jenis survei SFQ, FSQ, dan GFQ. Pada penelitian ini digunakan bahasa pemrograman *Python*. Alasan penulis dalam memilih bahasa pemrograman *Python* karena *Python* banyak didukung oleh *library* khususnya untuk melakukan *data mining*, memiliki banyak fitur *extension* yang dapat ditambahkan berdasarkan kebutuhan, serta merupakan bahasa yang mudah dipahami. Selain itu *Python* juga memiliki dukungan yang baik untuk pengguna karena memiliki komunitas sehingga dapat bermanfaat untuk mencari tahu tutorial penggunaan aplikasi bila mengalami kesulitan.

Kumpulan data OEQ yang telah terkumpul diolah dengan tahapan *data cleaning*, *data integration*, *data selection*, dan *data transformation*. Tahapan



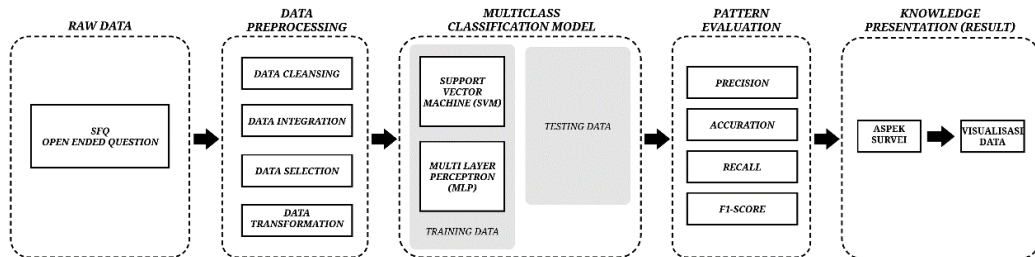
tersebut akan menghasilkan *train dataset* dan *test dataset*. Setelah itu *train dataset* dan *test dataset* diolah menggunakan teknik *classification*. Algoritma teknik *classification* yang digunakan adalah algoritma *Support Vector Machine* (SVM) dan *Multi Layer Perceptron* (MLP).

Kumpulan data CEQ yang telah terkumpul diolah dengan tahapan *data cleaning*, *data integration*, *data selection*, dan *data transformation*. Sebelum melakukan *clustering*, dilakukan penentuan jumlah *cluster* terbaik menggunakan *Elbow method*. Algoritma teknik *clustering* yang digunakan adalah *K-Means Clustering* dan *Fuzzy C-Means* (FCM).

Pengolahan data dengan setiap algoritma *classification* dan *clustering* tersebut menghasilkan model *classification* dan model *clustering*. Model *classification* label aspek OEQ pada survei SFQ berbeda dengan FSQ dan GFQ karena *classification* pelabelan aspek OEQ untuk SFQ dilakukan dengan *multi class classification*, sedangkan pelabelan aspek OEQ untuk FSQ dan GFQ dilakukan dengan *multi label Classification*. *Multi class classification* artinya pelabelan dilakukan dengan menetapkan masing-masing satu label pada setiap komentar. *Multi label classification* berarti pelabelan dilakukan dengan menetapkan satu atau lebih label pada setiap komentar. Pada *multi label classification*, sebuah komentar dapat memiliki satu atau lebih label aspek. Dalam menerapkan *multi label classification* pada algoritma SVM dibutuhkan *library* tambahan untuk mendukung *multi label classification* yang digunakan pada penelitian ini, yaitu *Classifier Chain* dan *One Versus Rest*. Algoritma MLP telah mendukung *multi label classification* sehingga tidak dibutuhkan tambahan *library multi label*.

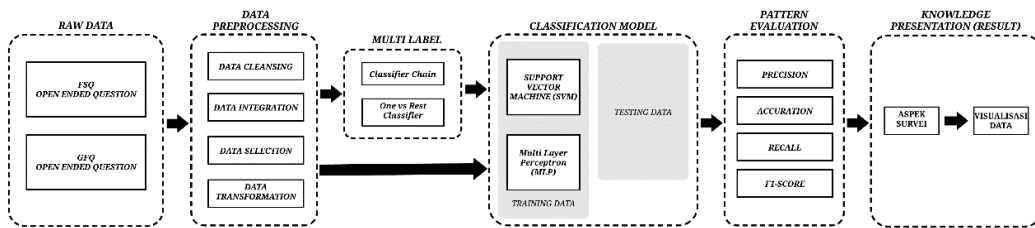
Setelah model *classification* dan model *clustering* dibuat, model *classification* akan dievaluasi dengan melakukan pengukuran *precision*, *accuration*, *recall*, dan *F1-Score*, sedangkan model *clustering* akan dievaluasi dengan menggunakan *silhouette coefficient*. Evaluasi tersebut bertujuan untuk menghasilkan perbandingan model terbaik sehingga mendapatkan hasil yang maksimal dalam mendapatkan label aspek komentar dan mendapatkan kelompok responden survei umpan balik. Pada penelitian ini juga akan dilakukan visualisasi data. Gambar 3.3 menunjukkan metodologi untuk melakukan analisis data survei

umpan balik berbentuk OEQ dengan jenis survei SFQ yang digunakan pada penelitian ini.



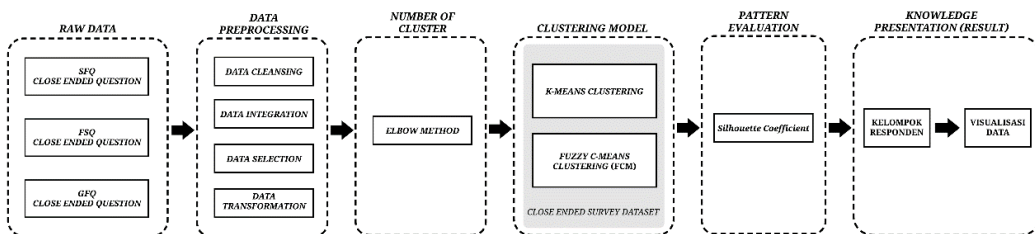
Gambar 3.3 Metodologi Survei SFQ bagian *Open Ended Question*

Gambar 3.4 menunjukkan metodologi untuk melakukan analisis data survei umpan balik berbentuk OEQ dengan jenis survei FSQ dan GFQ yang digunakan pada penelitian ini [8].



Gambar 3.4 Metodologi Survei FSQ dan GFQ bagian *Open Ended Question*

Gambar 3.5 menunjukkan metodologi untuk melakukan analisis data survei umpan balik berbentuk CEQ dengan jenis survei SFQ, FSQ, dan GFQ yang digunakan pada penelitian ini.



Gambar 3.5 Metodologi Survei SFQ, FSQ, dan GFQ bagian *Close Ended Question*

### 3.5 Perancangan Alur Menggunakan BPMN

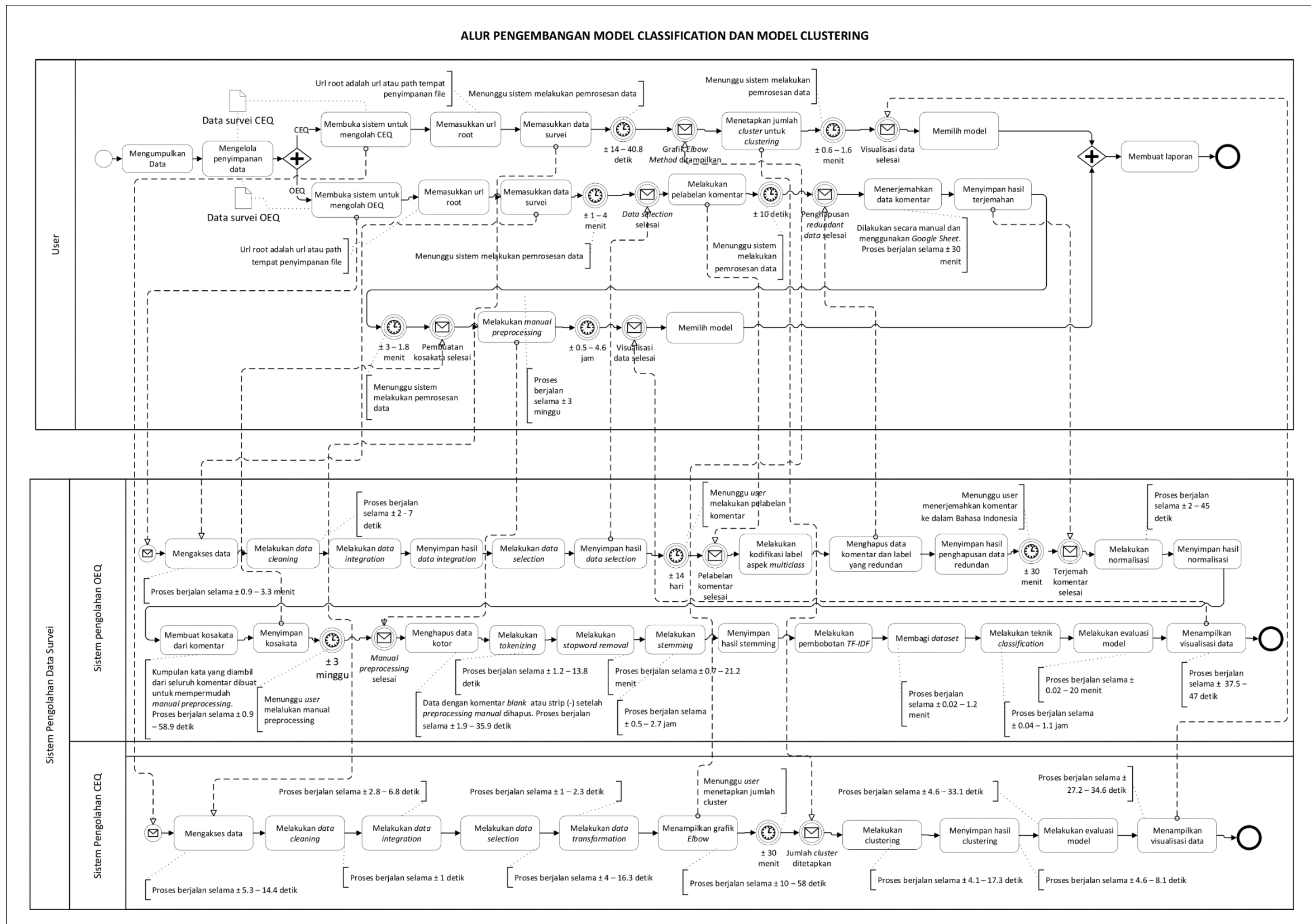
Pada bagian ini digambarkan alur sistem pengolahan data survei menggunakan BPMN. Penggambaran alur dalam penelitian ini dibagi menjadi dua, yaitu alur pengembangan model *classification* dan model *clustering*, dan alur pemanfaatan model *classification* untuk memprediksi label aspek survei berbentuk OEQ. Alur pengembangan model *classification* dan model *clustering* menggambarkan proses dalam mengembangkan model *classification* untuk bentuk survei OEQ dan model *clustering* untuk bentuk survei CEQ. Alur penggunaan atau pemanfaatan model *classification* menggambarkan proses dalam memanfaatkan model *classification* pada data komentar OEQ dalam memprediksi label aspek survei. Proses pengembangan model *classification* dan model *clustering* dimulai dari pengumpulan data hingga memilih model yang tepat sampai membuat laporan. Proses pemanfaatan model *classification* dimulai dari *user* membuka sistem demo pengolahan data survei hingga mendapatkan hasil pelabelan label aspek dan melakukan eksplorasi pengetahuan dari visualisasi data.

Model *classification* yang terbaik ditetapkan dari model yang memiliki nilai evaluasi terbaik. Model *classification* yang terbaik digunakan untuk menjadi model pembelajaran dalam memprediksi label aspek survei OEQ menggunakan data komentar baru. Data komentar baru yang dimaksudkan adalah data komentar OEQ yang belum memiliki label aspek survei. Oleh karena itu dikembangkan sistem untuk memanfaatkan model *classification* agar *user* dapat melakukan pemrosesan dan memprediksi label aspek pada data komentar baru.

Model *clustering* yang terbaik ditetapkan dari algoritma yang tepat serta penetapan jumlah *cluster* yang tepat untuk setiap data. Ketepatan tersebut dilihat berdasarkan nilai evaluasi yang terbaik. Model *clustering* yang didapatkan tidak digunakan dalam data baru karena penggunaan model *clustering* (pemilihan jumlah *cluster* dan pemilihan algoritma) mungkin tidak optimal pada kumpulan data yang lain. Model *clustering* terbaik pada penelitian ini akan digunakan untuk memperoleh data kelompok responden sehingga dapat dilakukan analisis lebih lanjut terkait hasil survei dalam bentuk CEQ.

### 3.5.1 Alur Pengembangan Model *Classification* dan *Clustering*

Proses pengembangan model digambarkan dalam dua *pool*, yaitu *pool User* dan *pool* Sistem Pengolahan Data Survei. *Pool User* merepresentasikan pengguna atau orang yang melakukan pengembangan model, sedangkan *pool* Sistem Pengolahan Data Survei merepresentasikan sistem yang dirancang untuk mengembangkan model menggunakan teknik *data mining* yaitu melakukan *preprocessing*, *classification*, *clustering* serta visualisasi data. *Pool* sistem pengolahan data survei dibagi menjadi dua *lane*, yaitu *lane* Sistem Pengolahan Data CEQ dan Sistem Pengolahan Data OEQ. Sistem pengolahan data CEQ menggambarkan bagian sistem yang menjalankan proses untuk mengembangkan model *clustering*, memperoleh hasil *clustering*, serta visualisasi data. Sistem pengolahan data OEQ menggambarkan sistem yang menjalankan proses dalam mengembangkan model *classification*. Model *classification* tersebut yang akan dimanfaatkan oleh *User* untuk mengembangkan sistem pemanfaatan model *classification* survei OEQ. Selain itu pada proses dalam sistem pengolahan data OEQ juga melakukan proses visualisasi data. Gambar 3.6 merupakan gambaran alur pengembangan model *classification* dan model *clustering*.



Gambar 3.6 BPMN Alur Pengembangan Model *Classification* & Model *Clustering*



Proses pengembangan model dimulai dari *User* yang mengumpulkan data kemudian mengelola penyimpanan data. Setelah mengelola penyimpanan data, *User* mulai melakukan pengembangan model *data mining* untuk survei CEQ dan OEQ. Pengembangan model untuk survei bentuk OEQ dilakukan dengan pemodelan *classification* yang dimulai dengan membuka sistem, kemudian memasukkan data survei ke dalam sistem dengan cara memberikan *path* lokasi data survei yang akan diolah. Kemudian Sistem Pengolahan Data OEQ melakukan pengaksesan data. Pengaksesan data dilakukan dalam durasi selama kurang lebih 0.9 menit sampai 3.3 menit dengan total 1.981 – 34.235 data untuk data SFQ, FSQ dan GFQ secara paralel. Setelah itu dilakukan *data cleaning*, lalu dilakukan *data integration* dan hasilnya disimpan di *file system*. Kemudian dilakukan *data selection*, kemudian *User* melakukan pelabelan label aspek secara manual. Pelabelan label aspek telah dilakukan pada penelitian sebelumnya, sehingga pelabelan label aspek pada penelitian ini dilakukan untuk menghindari kesalahan yang mungkin dilakukan pada penelitian sebelumnya. Setelah itu dilakukan kodifikasi label aspek *multi class*. Kodifikasi pelabelan aspek *multi label* tidak diperlukan dalam penelitian ini karena pelabelan telah dilakukan dengan cara kodifikasi. Setelah kodifikasi, dilakukan penghapusan data yang *redundan*, yaitu data yang memiliki komentar serta label aspek yang terduplikat. Setelah penghapusan *redundant data*, dilanjutkan dengan penyimpanan data ke *file system*. Kemudian *User* menerjemahkan komentar-komentar yang bukan menggunakan Bahasa Indonesia menjadi komentar Bahasa Indonesia dan menyimpan data hasil terjemahan tersebut. Setelah terjemahan selesai, sistem akan melakukan normalisasi dan menyimpan data hasil normalisasi tersebut. Kemudian membuat kosakata dari kumpulan komentar untuk mempermudah *User* melakukan *manual preprocessing*. Setelah itu, *User* melakukan *manual preprocessing* selama kurang lebih 3 minggu. Setelah *manual preprocessing* selesai, sistem melakukan *tokenizing*, *stopword removal*, *stemming*, pembobotan TF-IDF dan pembagian *dataset* secara berurutan. Kemudian dilakukan teknik *classification* dengan model yang telah dirancang. Setelah *classification*, dilakukan evaluasi model dan visualisasi data. Visualisasi data memungkinkan *User* untuk dapat melihat *insight*



selama pembuatan model termasuk laporan pengolahan data dan hasil evaluasi model yang diperoleh, sehingga *User* dapat memilih model terbaik berdasarkan hasil evaluasi tersebut.

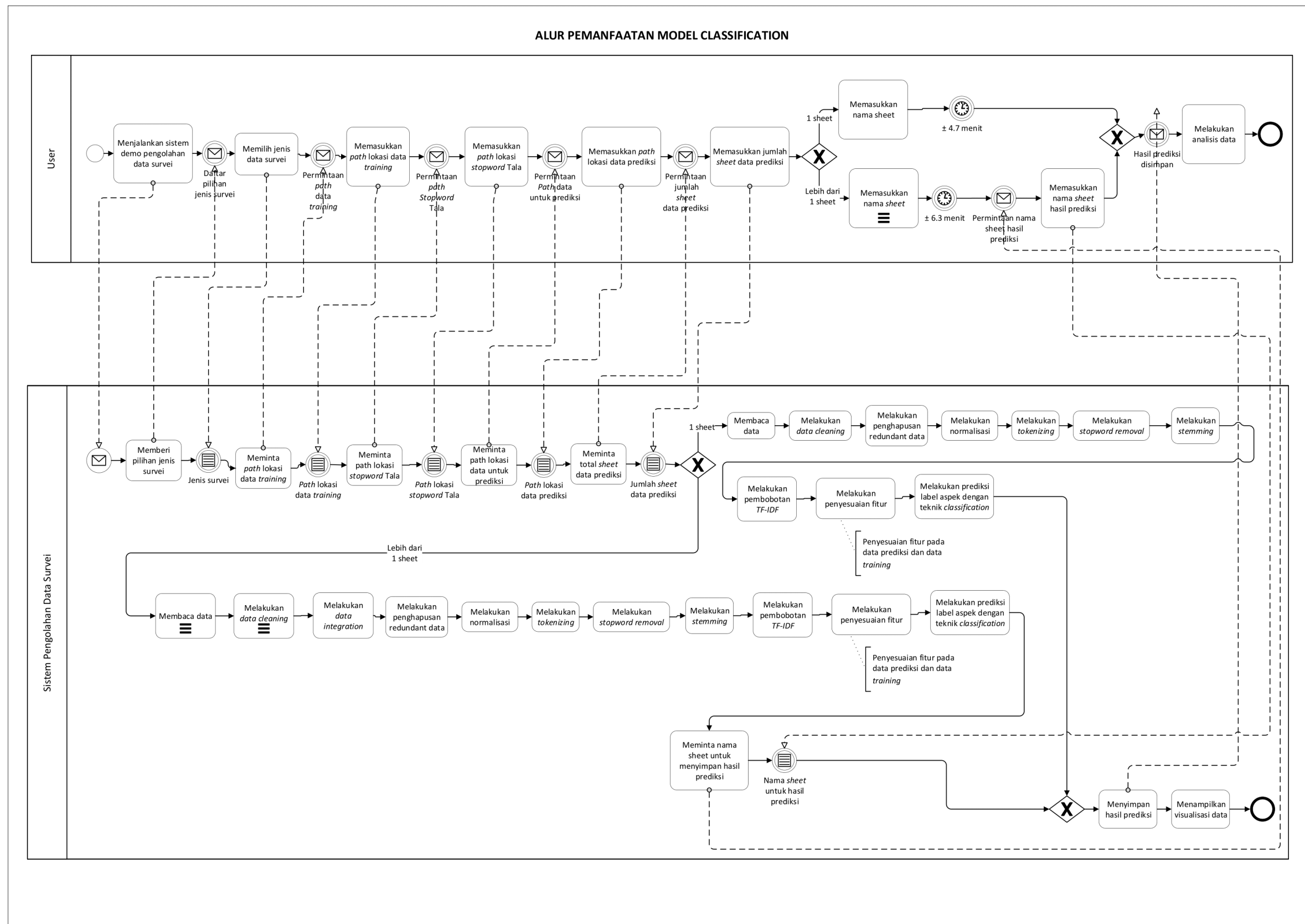
Pengembangan model untuk CEQ dilakukan dengan pemodelan *clustering* yang dimulai dengan membuka sistem kemudian memasukkan data survei ke dalam sistem dengan cara memberikan *path* lokasi data survei yang akan diolah. Kemudian sistem pengolahan data CEQ melakukan pengaksesan data. Pengaksesan data dilakukan dalam durasi selama kurang lebih 5.3 detik sampai 14.4 detik dengan total 4.330 – 6.677 data untuk data SFQ, FSQ dan GFQ secara paralel. Setelah itu dilakukan *data cleaning*, *data integration*, *data selection*, *data transformation* dan menampilkan *Elbow plot*. Berdasarkan *Elbow plot*, *User* menentukan jumlah *cluster* untuk menghasilkan hasil *clustering* yang optimal. Setelah jumlah *cluster* ditetapkan, sistem melakukan *clustering* dengan algoritma dan jumlah *cluster* yang telah ditetapkan oleh *User*. Kemudian didapatkan hasil *clustering* berupa data responden yang telah dikelompokkan. Hasil *clustering* tersebut disimpan dan dilakukan evaluasi model *clustering*. Setelah itu, sistem juga melakukan visualisasi data termasuk perbandingan evaluasi satu model *clustering* dengan model *clustering* yang lain. Setelah visualisasi data, *User* menentukan model *clustering* yang terbaik dalam mendapatkan kelompok responden survei CEQ untuk dilakukan analisis lebih lanjut. Setelah user melakukan pemilihan model *classification* dan model *clustering* dengan melihat hasil evaluasi masing-masing model, *user* membuat laporan yang dimasukkan ke dalam laporan dan analisis penelitian.

### **3.5.2 Alur Pemanfaatan Model Classification**

Proses pemanfaatan model *classification* digambarkan dalam dua *pool*, yaitu *pool User* dan sistem pengolahan data survei. *Pool User* merepresentasikan orang atau pengguna yang menggunakan sistem pemanfaatan model *classification*. Sistem pemanfaatan model *classification* digunakan untuk memprediksi label aspek untuk data survei OEQ yang belum memiliki label aspek. *Pool* sistem pengolahan data pada skema BPMN ini menggambarkan bagian sistem yang menjalankan proses untuk memprediksi label aspek pada komentar dalam survei OEQ yang

belum memiliki label aspek. Gambar 3.7 merupakan gambaran alur pemanfaatan model *classification* untuk memprediksi label aspek survei umpan balik OEQ.





Gambar 3.7 BPMN Alur Pemanfaatan Model Classification

Proses pemanfaatan model *classification* untuk memprediksi label aspek survei dimulai dari *User* yang menjalankan sistem demo pengolahan data survei. *User* diberi pilihan oleh sistem untuk memasukkan jenis survei yang akan diolah untuk memprediksi label aspeknya. Setelah *User* memasukkan jenis survei yang akan diolah, sistem meminta masukan untuk *path* lokasi data-data yang dibutuhkan untuk prediksi. Salah satu data tersebut adalah *data training* yang digunakan pada pembuatan model. *Data training* tersebut adalah data yang telah diolah dan dijadikan sebagai masukan pembelajaran model pada saat pengembangan model. Selain itu, data yang dibutuhkan adalah data *stopword* Tala yang digunakan pada saat pembuatan model. *Data stopwords* ini digunakan pada proses selanjutnya dalam tahap *stopword removal*. Setelah memasukkan *path* lokasi data *stopword* Tala, sistem meminta *User* untuk memasukkan *path* lokasi data survei umpan balik OEQ yang belum memiliki label aspek untuk diprediksi. Setelah *User* memasukkan *path* lokasi data OEQ untuk diprediksi, sistem meminta jumlah *sheet Microsoft Excel* yang diinginkan *User* untuk dilakukan prediksi label aspek. Setelah *User* memasukkan total *sheet*, proses selanjutnya dibagi menjadi dua alur, yaitu proses dengan sebuah jumlah *sheet* dan proses dengan *sheet* yang berjumlah lebih dari satu. Sistem meminta *User* untuk memasukkan nama *sheet Microsoft Excel* sebanyak total *sheet* yang ditentukan oleh *User*. Setelah *User* memasukkan nama *User*, sistem melakukan pembacaan data dan *data cleaning* pada data untuk setiap *sheet* menyesuaikan jumlah *sheet* yang *User* tentukan. Setelah *data cleaning*, sistem melakukan *data integration* jika jumlah *sheet* yang ditentukan *User* berjumlah lebih dari satu *sheet*. Tahap *data integration* akan diabaikan pada pengolahan data berjumlah satu *sheet*. Selanjutnya sistem akan melakukan penghapusan data yang terduplikat, normalisasi, *tokenizing*, *stopword removal*, *stemming*, serta pembobotan TF-IDF secara berurutan. Setelah pembobotan TF-IDF, sistem pengolahan data survei untuk memprediksi label aspek menjalankan fungsi untuk menyesuaikan fitur data, yaitu pembobotan kata-kata pembentuk komentar yang berbentuk *uni-bigram*. Fitur *uni-bigram* pada data baru disesuaikan dengan fitur *uni-bigram* pada *data training* agar dapat dilakukan *classification*. Setelah melakukan penyesuaian fitur, sistem melakukan prediksi label aspek dengan teknik

*classification* dengan model terbaik yang telah dihasilkan dan dipilih pada saat pengembangan model. Setelah melakukan prediksi dengan teknik *classification*, jika jumlah *sheet* yang ditetapkan *User* lebih dari satu, maka sistem akan meminta *User* untuk memasukkan nama *sheet* untuk menyimpan hasil prediksi label aspek survei umpan balik. Setelah *User* memasukkan nama *sheet* untuk hasil prediksi, sistem akan menyimpan hasil prediksi dan melakukan visualisasi data. Setelah visualisasi data ditampilkan, *User* dapat melakukan eksplorasi pengetahuan terhadap data yang diprediksi melalui visualisasi data yang interaktif. Visualisasi data interaktif yang dimaksudkan yaitu visualisasi data yang menampilkan informasi sesuai dengan perintah *User*. Visualisasi interaktif ini tidak memberikan seluruh informasi dalam satu waktu, namun *User* dapat memilih informasi mana saja yang ingin ditampilkan.

