

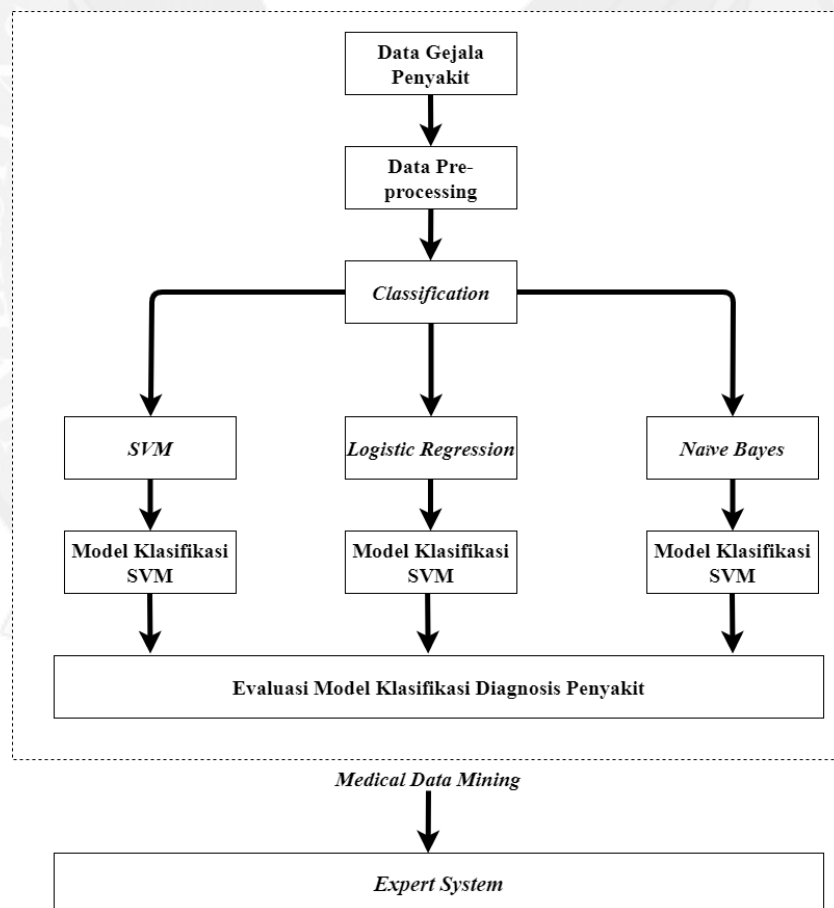
## BAB III

### METODOLOGI PENELITIAN

Bab ini membahas tentang kerangka teoritis yang menjadi dasar penelitian, serta metode – metode yang dipakai untuk menjawab pertanyaan penelitian. Bab ini terdiri dari kerangka teoritis, tahapan penelitian, metode pengumpulan data, dan metode analisis data.

#### 3.1 *Theoretical Framework*

Bagian ini menjelaskan kerangka teoritis yang digunakan sebagai dasar dalam melakukan penelitian. Teori – teori yang dipakai dapat dilihat pada Gambar 3.1.



Gambar 3.1 Theoretical Framework

Gambar 3.1 menunjukkan bahwa teori *data mining* (Han, Kamber, & Pei, 2012) memiliki beberapa proses yang perlu dilakukan untuk menghasilkan

pengetahuan. Salah satu bidang dimana *data mining* dipakai untuk menghasilkan pengetahuan adalah bidang kesehatan atau medis. Pada bidang medis, *data mining* banyak digunakan untuk berbagai hal, salah satunya adalah memprediksi suatu penyakit. Data penyakit tersebut dapat memiliki berbagai bentuk, seperti *medical transcription*, tes laboratorium, sinyal frekuensi dari alat medis, *electronic health record*, dan video. Data yang digunakan pada penelitian ini adalah *medical transcription* berupa gejala – gejala penyakit dan hasil tes laboratorium.

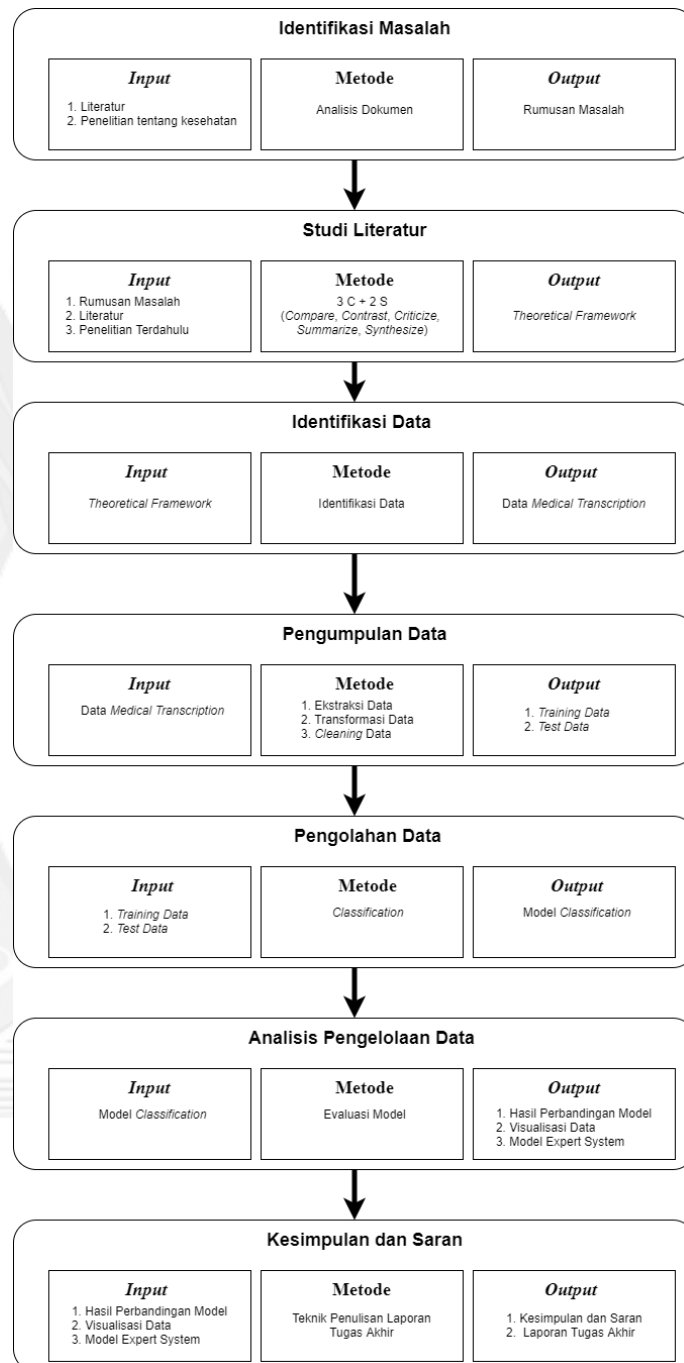
Berdasarkan penelitian-penelitian yang dilakukan oleh K. Tomita, R. Nagao, H. Touge, T. Ikeuchi, H. Sano, A. Yamasaki, dan Y. Tohda pada tahun 2017, A. K. Arslan, C. Colak, dan M. E. Sarihan pada tahun 2016, A. Khemphila dan V. Boonjing pada tahun 2010, R. Alizadehsani, J. Habibi, M. J. Hosseini, H. Mashayekhi, R. Boghrati, A. Ghandeharioun, B. Bahadorian, dan Z. A. Sani pada tahun 2013, K. Eyasu, W. Jimma, dan T. Tadesse pada tahun 2020, S. Malik, R. Khadgawat, S. Anand, dan S. Gupta pada tahun 2016, L. Tapak, H. Mahjub, O. Hamidi, dan J. Poorolajal pada tahun 2013, I. Al-Turaiki, M. Alshahrani, dan T. Almutairi pada tahun 2016, C. Lin, E. W. Karlson, H. Canhao, T. A. Miller, D. Dligach, P. J. Chen, R. N. G. Perez, Y. Shen, M. E. Weinblatt, N. A. Shadick, R. M. Plenge, dan G. K. Savova pada tahun 2013, dan C. A. Turner, A. D. Jacobs, C. K. Marques, J. C. Oates, D. L. Kamen, P. E. Anderson, dan J. S. Obeid pada tahun 2017, teknik *data mining* digunakan dalam bidang medis untuk menganalisis data – data medis sehingga dapat memprediksi keberadaan atau kejadian dari suatu penyakit. Pada penelitian – penelitian tersebut, teknik *data mining* yang digunakan adalah teknik klasifikasi dengan berbagai macam algoritma. Pada penelitian ini, teknik *data mining* digunakan untuk melakukan klasifikasi penyakit pada *medical transcription* dengan menggunakan algoritma *Support Vector Machine* (SVM), *Logistic Regression*, dan *Naïve Bayes*. Penelitian ini berfokus membandingkan performa dari algoritma teknik klasifikasi untuk melakukan klasifikasi penyakit. Model yang didapatkan dari pengelolaan data *medical transcription* dengan teknik klasifikasi akan dievaluasi dengan menghitung akurasi, *error rate*, sensitivitas (*recall*), presisi (*precision*), dan *F-score* dari setiap model untuk mendapatkan model terbaik untuk melakukan klasifikasi penyakit.

### 3.2 Tahapan Penelitian

Tahapan penelitian menjelaskan proses – proses yang dilakukan saat melakukan penelitian. Dalam bagian ini terdapat alur penelitian berupa *input*, metode, dan *output* pada setiap fase alur penelitian yang ditunjukkan pada gambar 3.2.

Pada Gambar 3.2 dapat dilihat bahwa tahapan dimulai dengan tidak banyak ditemukannya penelitian *data mining* yang memprediksi lebih dari satu penyakit. Pembuatan sistem pakar (*expert system*) dinilai mampu menghasilkan model *data mining* yang memprediksi lebih dari satu penyakit menghasilkan sebuah rumusan masalah, yaitu: “Bagaimana cara merancang dan mengembangkan sebuah sistem pakar menggunakan metode klasifikasi *data mining* yang dapat memprediksi penyakit berdasarkan gejala – gejala yang diberikan?” Oleh sebab itu, tahapan penelitian dilakukan untuk menjawab pertanyaan penelitian. Tahapan selanjutnya adalah pemilihan teknik yang dapat digunakan untuk melakukan teknik klasifikasi yang diperoleh dari studi literatur dan penelitian sebelumnya. Semua tahapan tersebut menghasilkan sebuah *theoretical framework* untuk penelitian ini. Setelah mendapatkan *theoretical framework*, langkah selanjutnya adalah melakukan identifikasi terhadap data yang akan digunakan untuk teknik klasifikasi. Hasil identifikasi tersebut diperoleh data *medical transcription* untuk melakukan teknik klasifikasi. Tahapan selanjutnya adalah pengelolaan data dengan melakukan tahap *data cleaning*, *data integration*, *data selection*, *data transformation*, dan *data mining* untuk mendapatkan *training data* dan *test data*. *Training data* dan *test data* digunakan untuk memperoleh model klasifikasi terhadap data *medical transcription*. *Training data* digunakan untuk melatih dan menghasilkan model klasifikasi terhadap data *medical transcription* untuk memprediksi penyakit yang ditentukan. *Test data* digunakan untuk menguji hasil dari model klasifikasi yang sudah dilatih. Setelah itu, model dievaluasi dengan memeriksa akurasi, *error rate*, sensitivitas, presisi dan *F<sub>1</sub> score*. Hasil dari evaluasi dapat menghasilkan model terbaik untuk memprediksi penyakit berdasarkan *medical transcription*. Model sistem pakar akan dibuat berdasarkan hasil evaluasi dari algoritma. Model sistem pakar dibuat menggunakan bahasa pemrograman *Python*. Kesimpulan dan saran

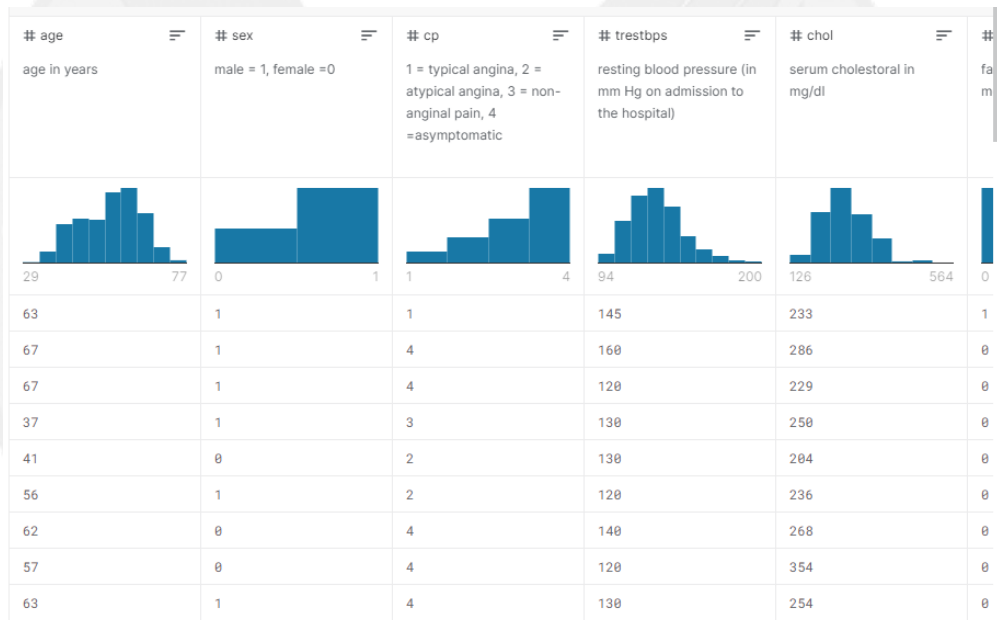
penelitian diambil setelah data - data diproses dan hasil evaluasi dari perbandingan setiap algoritma klasifikasi.



Gambar 3.2 Tahapan Penelitian

### 3.3 Metode Pengumpulan Data

Data yang digunakan pada penelitian ini diambil melalui *platform open-source* yang menyediakan berbagai macam *dataset* seperti *Kaggle* dan *UCI Machine Learning Repository*. Dataset yang digunakan berupa *medical transcription* digital. *Medical transcription* merupakan sebuah catatan medis yang biasanya berisi data pasien seperti umur, jenis kelamin, dan gejala – gejala penyakit yang dialami. Gambar 3.3 menunjukkan contoh dari *medical transcription*. Dataset memiliki *extension* .csv dan .data. Biasanya terdapat keterangan dari dataset dalam bentuk deskripsi dalam *website* atau file berekstensi .names yang dapat dibuka menggunakan program – program seperti *Notepad*.



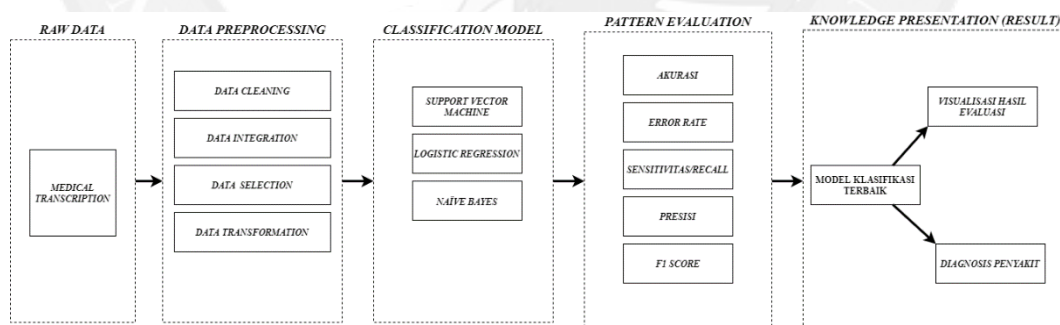
Gambar 3.3 Contoh *Medical Transcription* Digital  
Sumber : Kaggle.com

### 3.4 Metode Analisis Data

Penelitian ini menggunakan teknik klasifikasi untuk mendapatkan model label diagnosis penyakit. Penelitian ini akan dikembangkan dengan menggunakan bahasa pemrograman *Python*. Penelitian ini menggunakan bahasa pemrograman *Python* karena mendukung mendukung banyak metode dan algoritma untuk melakukan berbagai teknik *data mining*, memiliki banyak *library open-source* dari

algoritma *data mining*, dan memiliki fitur menampilkan hasil analisis data secara langsung. Selain itu, *Python* merupakan bahasa pemrograman paling sesuai untuk penelitian ini karena mendukung banyak metode dan algoritma *data mining*. Selain itu, aplikasi *Python* memiliki dukungan yang baik untuk pengguna melalui komunitas yang ada, sehingga dapat berguna bagi penelitian ini untuk mendapatkan tutorial penggunaan aplikasi apabila mendapatkan kesulitan.

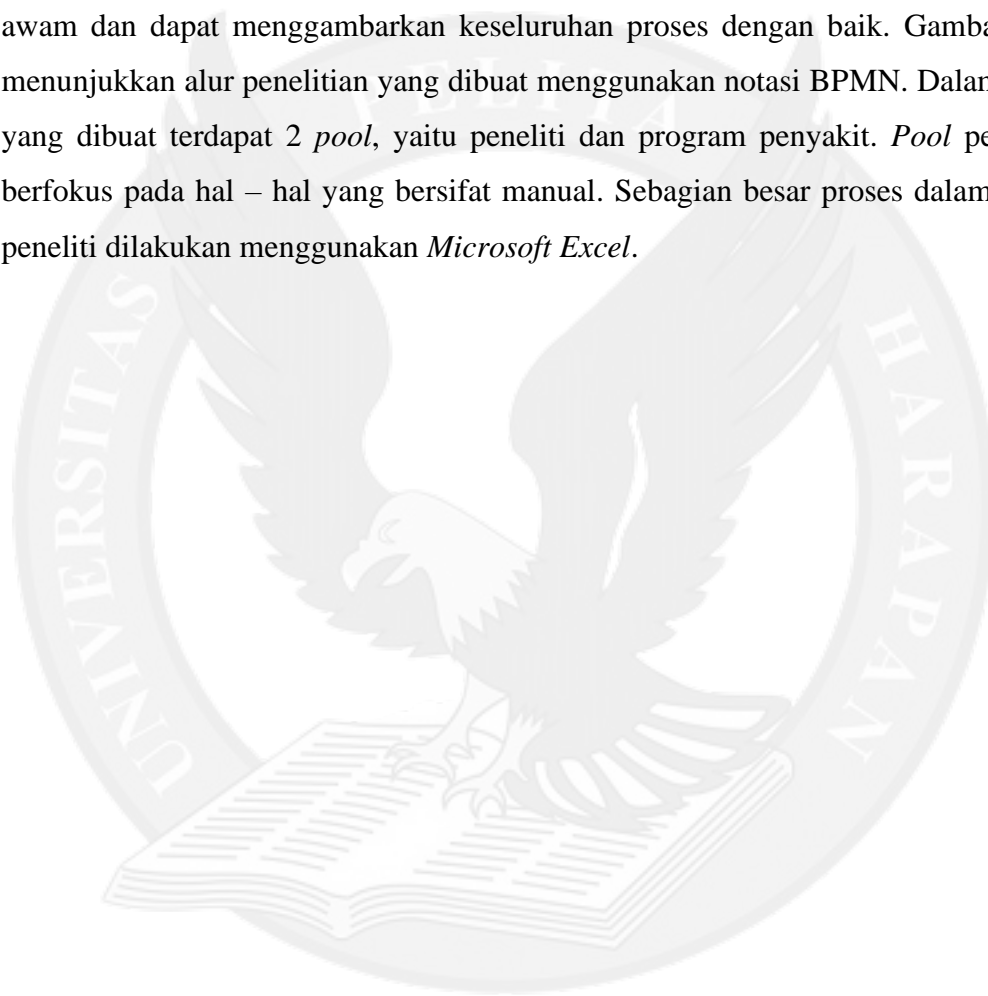
Data yang sudah terkumpul akan dilakukan pengolahan data dengan tahapan *data cleaning*, *data integration*, *data selection*, dan *data transformation*. Hasil dari tahapan-tahapan tersebut adalah kumpulan data siap pakai untuk dibagi menjadi *training data* dan *test data* pada tahap *data mining*. Setelah itu, *training data* dan *test data* akan diolah menggunakan algoritma SVM, *Logistic Regression*, dan *Naïve Bayes* untuk mendapatkan model klasifikasi. Gambar 3.4 menunjukkan metodologi untuk analisis data *medical transcription* yang digunakan pada penelitian ini. Setelah model klasifikasi sudah dibuat, selanjutnya model akan dievaluasi dengan melakukan pengukuran akurasi, *error rate*, sensitivitas, presisi (*precision*), dan *F1 score*. Evaluasi ini dilakukan sehingga dapat menunjukkan algoritma dengan performa yang paling terbaik. Algoritma dengan hasil evaluasi paling baik akan digunakan untuk mendapatkan diagnosis penyakit yang paling tepat dan cepat.

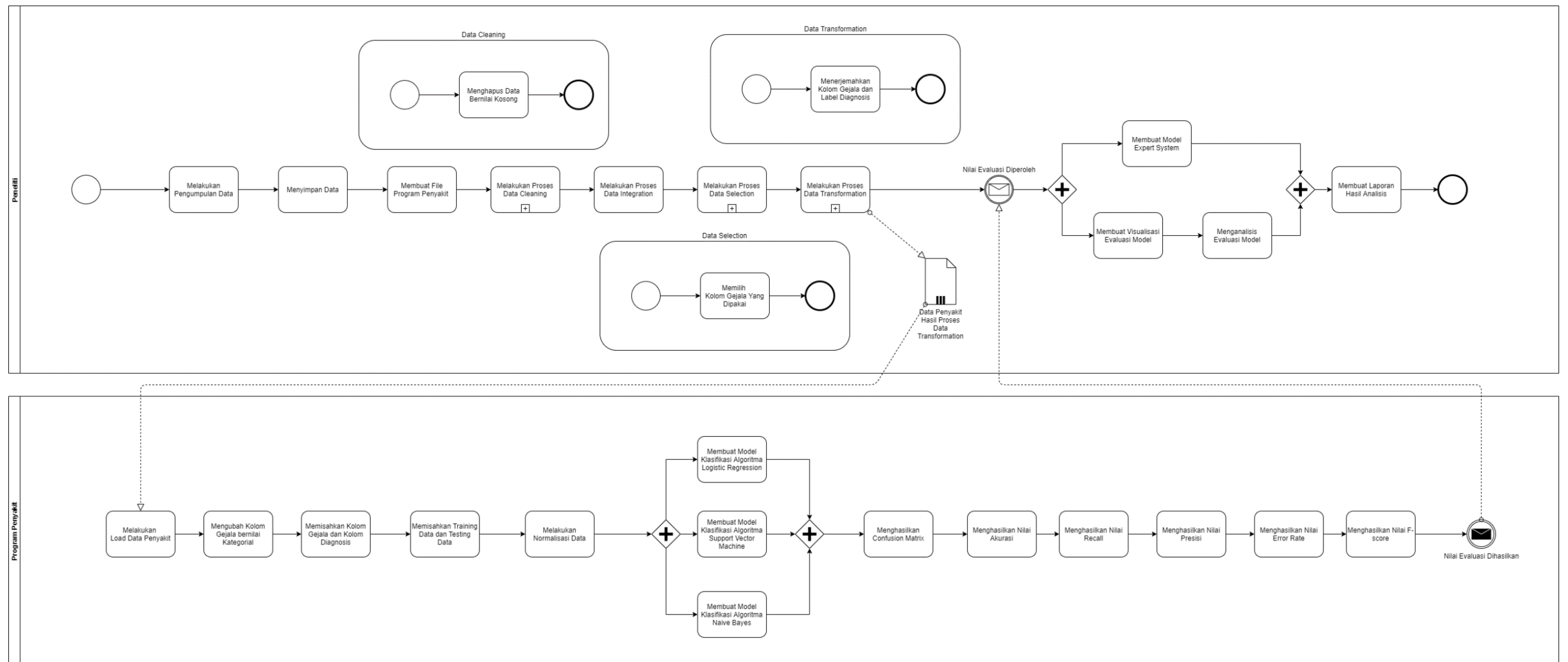


Gambar 3.4 Metode Analisis Data

### 3.5 Alur Penelitian

Alur penelitian dibuat untuk menunjukkan proses penelitian dalam melakukan pengolahan data sehingga menjadi diagnosis yang dihasilkan dalam program penyakit dan juga nilai evaluasi yang dihasilkan. Alur penelitian ini dibuat menggunakan bahasa *modeling* BPMN. BPMN dipilih untuk mendeskripsikan alur penelitian dikarenakan memiliki notasi – notasi yang mudah untuk dipahami orang awam dan dapat menggambarkan keseluruhan proses dengan baik. Gambar 3.5 menunjukkan alur penelitian yang dibuat menggunakan notasi BPMN. Dalam alur yang dibuat terdapat 2 *pool*, yaitu peneliti dan program penyakit. *Pool* peneliti berfokus pada hal – hal yang bersifat manual. Sebagian besar proses dalam *pool* peneliti dilakukan menggunakan *Microsoft Excel*.



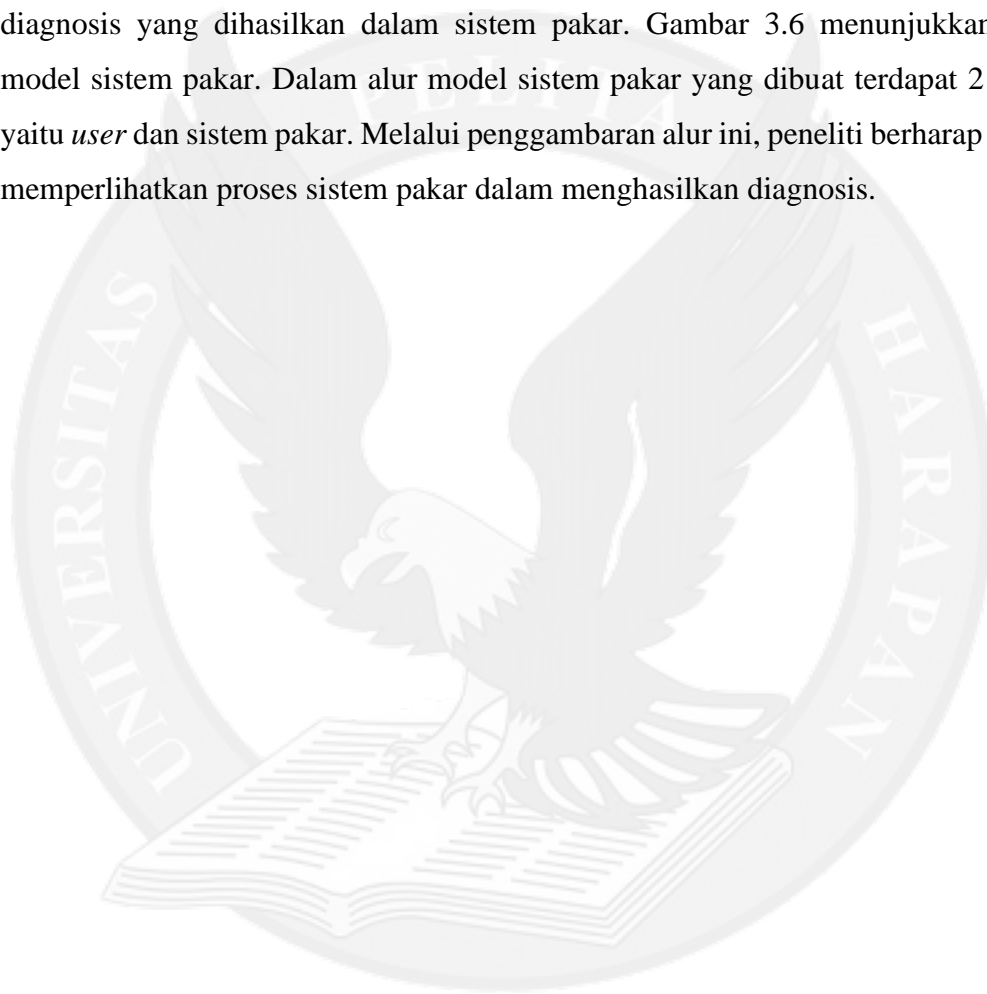


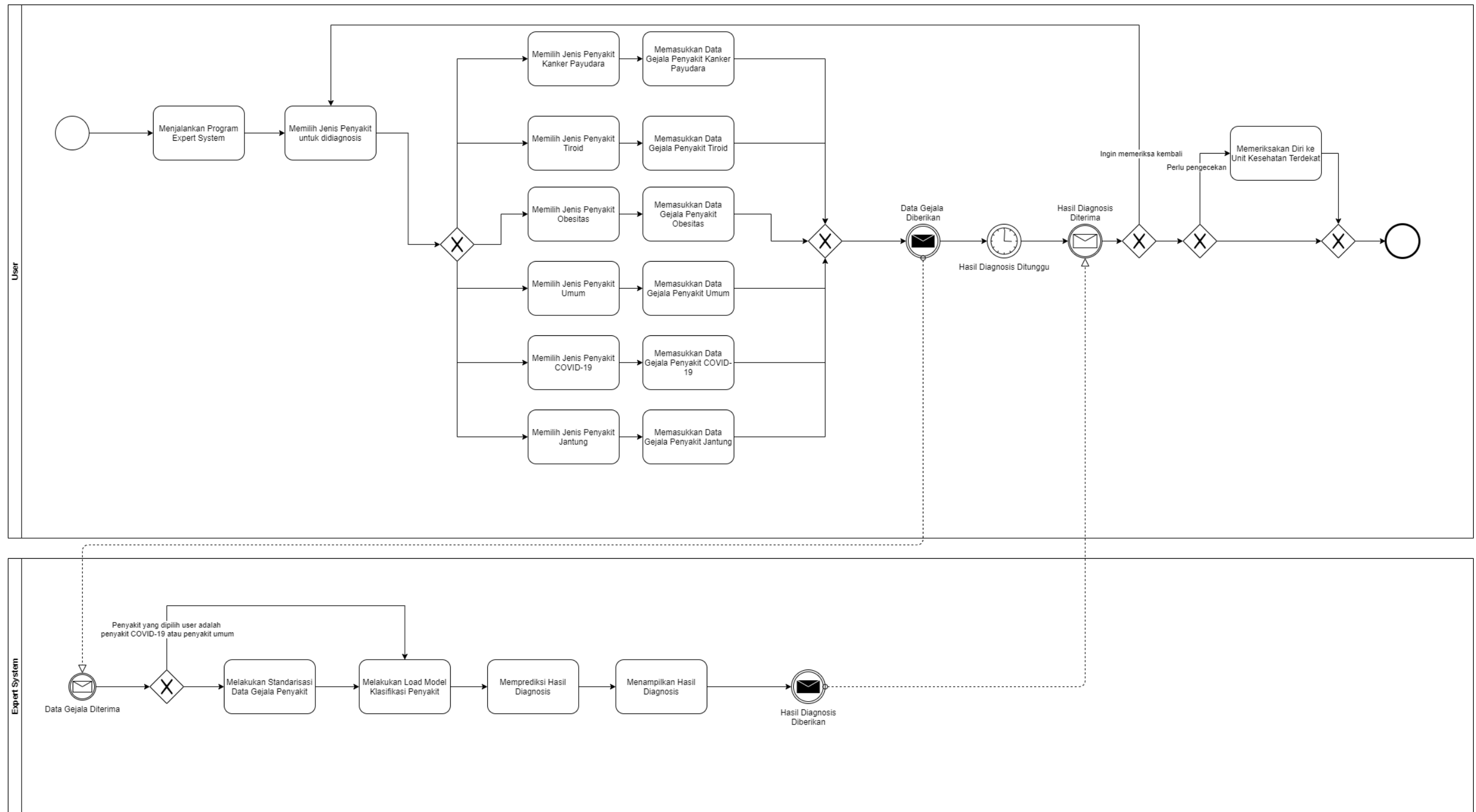
Gambar 3.5 Alur Penelitian



### 3.6 Alur Model Sistem Pakar

Alur model sistem pakar (*expert system*) dibuat untuk menunjukkan bagaimana alur dari user memilih jenis layanan hingga menghasilkan diagnosis sesuai tipe penyakit yang dipilih. Alur model sistem pakar digambarkan menggunakan bahasa *modeling* BPMN. BPMN digunakan untuk mendeskripsikan alur bagaimana penelitian ini melakukan pengolahan data sehingga menjadi diagnosis yang dihasilkan dalam sistem pakar. Gambar 3.6 menunjukkan alur model sistem pakar. Dalam alur model sistem pakar yang dibuat terdapat 2 *pool*, yaitu *user* dan sistem pakar. Melalui penggambaran alur ini, peneliti berharap dapat memperlihatkan proses sistem pakar dalam menghasilkan diagnosis.





Gambar 3.6 Alur Model Sistem Pakar