Expert System That Detects COVID-19 Using Forward Chaining Algorithm


Kata kunci: Covid-19, virus corona, sistem pakar, forward chaining.

Abstract. The spread of Covid-19 disease since early year 2020 which caused by coronavirus has made the world in great pandemic. Besides the vaccine has not been discovered yet, it is also the cost for doing rapid test or polymerase chain reaction (PCR) are expensive. By using the information technology that is expert system, so the aim of this research is to develop expert system to detect Covid-19 disease based on the confirmed symptoms of Covid-19. This expert system will be using the algorithm of forward chaining.

Keywords: Covid-19, coronavirus, expert system, forward chaining.

1. Introduction
In early year 2020 the world was shake by the outbreak of new type of coronavirus (SARS-CoV-2) and the disease called Corona Virus Disease 2019 (Covid-19). It is known this virus origin was from Wuhan city in China in late year 2019 [1]. Until now coronavirus has spread in more than 200 countries in the world. The total of infected people on May, 16 2020 has reached 4,518,074, and the total people died is 307,825 worldwide [2].

Coronavirus is spreading around the world, but there are still no vaccines to protect the body against the disease it causes [3]. To control the spread of this disease, some big cities in some countries which is become the epicentral of the spread is being locked down. While in Indonesia is manage with high scale social distancing which is called PSBB.

In response to the growing COVID-19 pandemic and shortages of laboratory-based molecular testing capacity and reagents, multiple diagnostic test manufacturers have developed and begun selling rapid and easy-to-use devices to facilitate testing outside of laboratory settings. One type of rapid diagnostic test (RDT) detects the presence of viral proteins (antigens) expressed by the COVID-19 virus in a sample from the respiratory tract of a person [4]. Half or more of COVID-19 infected patients might be missed by such test, depending on the group of patients tested. These assumptions urgently require further study to understand whether they are accurate. Additionally, false-positive results – that is, a test showing that a person is infected when they are not – could occur if the antibodies on the test strip also recognize antigens of viruses other than COVID-19 [4]. However, the SARS-CoV-2 RNA was detected in only 32% of oropharyngeal swabs, which was significantly lower than the level in nasal swabs (63%) [5].

The Indonesian government suggest to do RDT, but because of lacked of kits to do the test and the cost of the test quite expensive which is around 1.5 K in IDR, so the test is prioritized only to those people with higher risk to get COVID-19 [6]. In order to help people to detect whether they are infected by coronavirus or not, the aim of this research is to develop web-based application which worked as expert system using the algorithm of forward chaining.

2. Theoretical Framework and Methodology
2.1. Expert System
According to F. Bobillo in al Ajlan, expert system is programs that gives advice automatically that trying to mimic processes of thinking and knowledge from experts to reach goal from certain problem [7].
For the last decades, expert system has become the main practice application from AI research. Nowadays, there are many systems used in almost every operational field around the world. From simple gadget like handphone until robotics in manufacture industry and medic [8].

2.2. **Forward Chaining**
Forward-chaining is a general concept of thought that controlled by data (data-driven) that is, thoughts that focus on known data or facts. Forward-chaining can be used within an agent to produce conclusion from incoming perceives, usually without specific query.

Forward-chaining starts from available data and using rules of inferences to get another data until the goal or conclusion is reached. Inference mechanism that utilizes forward-chaining search for rules of inferences until one of antecedent (IF-THEN clause) is true. When a rule is found, then the inference mechanism make decision or consequent (THEN clause), that generate new information from available data. The mechanism will iterate through this process until the goal is reached [8].

2.3. **Backward Chaining**
Backward chaining is the opposite of forward chaining. Backward chaining starts from goals and works backward to see if there is data which will allow to it to conclude ant of these goals [9]. Sometimes forward chaining used together with backward chaining to get optimal solution.

2.4. **Inference Engine**
The inference engine is one of main components of expert system that influences the performance of expert system. The task of inference engine is to give answers and reasons to users by inference the knowledge of expert system [10].

2.5. **Methodology**
This research is an applied research. The writer tries to implement the forward-chaining algorithm within a study case that is; how to detect COVID-19 disease. So here is the framework that conducted to do the research:

![Research Framework](image)

3. **Result and Discussion**

3.1. **Gathering Knowledge Base (KB)**
Expert system is a system that adopted from knowledges that belong to an expert in certain field. These knowledges are used as knowledge base. The writer has gathered data about coronavirus and its disease which is called COVID-19. These data are about symptoms of person who infected by coronavirus.

According to Yuliana in her journal Corona Virus Disease (Covid-19) [1], mentioned symptoms of person who indicated infected by coronavirus, grouped as follows:

1. Patient in Observation (called PDP)
   a. Someone who has:
      i. Fever >= 38° Celsius
ii. Cough or flu or sore throat.
iii. Pneumonia mild or heavy according to clinical and/or image of radiology.
   (patient with immunocompromised percent of probability is atypical) and comes with one of these condition as follow:
   1. Having historical trip to China or region or country where the disease spread within 14 days before the first symptom.
   2. Medical officer who get sick with same symptom after nursing patient with acute respiratory infection with unknown caused, without bothered by historical trip or live.

b. Patient with acute respiratory infection from mild to heavy and one of these within 14 days before symptom:
   i. Contacted with patient confirmed or probable Covid-19, or
   ii. Having historical contact with infected animal (if its has been identified), or
   iii. Working or visiting public health service which has cases of confirmed or probable COVID-19 in China or region or country that has been infected.
   iv. Having historical trip to Wuhan and having fever >= 38°C Celsius or having historical fever.

2. Person in Observation (called ODP)
   Someone who has fever symptom or historical fever without pneumonia which has historical trip to China or region or country that has been infected, and not having one or more of historical exposure as follow:
   a. Contacted with patient confirmed Covid-19
   b. Working or visiting public health service which has cases of confirmed or probable COVID-19 in China or region or country that has been infected.
   c. Having historical contact with infected animal (if it has been identified) in China or region or country that has been infected.

Meanwhile CDC has difference symptoms which is as mention below:
1. Cough
2. Shortness of breath
3. Or at least two of these symptoms:
   a. Chills
   b. Muscle pain
   c. Repeated shaking
   d. Loss of taste or smell (point iv is according to CDC) [11].

3.2. Rules Construction
From knowledge base above, we need to simplify in a form of propositional logic, so we can construct rules that easier to be understood by inference mechanism. Here is the simplification that match propositional logic form. For acknowledge the rules constructed here is only for patient in observation (PDP) and the symptoms only from Yuliana’s data.

A. PDP
B. Fever >= 38°C Celsius
C. Cough
D. Flu
E. Sore throat
F. mild or heavy according to clinical and/or image of radiology
G. Immunocompromised
H. Having historical trip to China or region or country where the disease spread within 14 days before the first symptom.
I. Medical officer who get sick with same symptom after nursing patient with acute respiratory infection with unknown caused, without bothered by historical trip or live.
J. Patient with acute respiratory infection from mild to heavy.
K. Contacted with patient confirmed or probable Covid-19 within 14 days before symptom.
L. Having historical contact with infected animal (if it has been identified) within 14 days before symptom.
M. Working or visiting public health service which has cases of confirmed or probable COVID-19 in China or region or country that has been infected within 14 days before symptom.

Rules:
R1. IF B AND (C OR D OR E) AND F THEN A
R2. IF B AND (C OR D OR E) AND G AND (H OR I) THEN A
R3. IF J AND (K OR L OR M OR (H AND B)) THEN A

Rules up above needs to be simplified furthermore in order to match to the rules of knowledge base that only contains certain clauses with positive value.

Rules R1 will be elaborated using propositional logic notation (AND = \( \land \), OR = \( \lor \), THEN = \( \Rightarrow \), NOT = \( \neg \)) as follow:

\[
\begin{align*}
S & \Rightarrow A \\
R \land F & \Rightarrow S \\
B \land Q & \Rightarrow R \\
P \lor E & \Rightarrow Q \\
C \lor D & \Rightarrow P
\end{align*}
\]

We are adding new literals that is; P, Q, R, S as conclusions literals. For R2 we will do the same as above.

\[
\begin{align*}
U \land T & \Rightarrow S \\
R \land G & \Rightarrow U \\
H \lor I & \Rightarrow T
\end{align*}
\]

We are adding new literals that is; U and T as conclusions literals. And R3 is as follow:

\[
\begin{align*}
X \lor V & \Rightarrow S \\
W \lor M & \Rightarrow X \\
K \lor L & \Rightarrow W \\
H \land B & \Rightarrow V
\end{align*}
\]

Those rules are becoming our knowledge base, and we will store it in a text file. The program will read it first before executing new data from the user input.

3.3. Implementation

3.3.1. Use Case Diagram Design

![Use Case Diagram](image)

Figure 2. Use Case Diagram

3.3.2. Sequence Diagram Design
3.3.3. Database Design

```plaintext
# dm Tables

## knowledge_base
- `id`: INT
- `literal`: VARCHAR(50)
- `sentence`: TEXT
- `PK_knowledge_base`: INT

## participants
- `id`: INT
- `nama`: VARCHAR(50)
- `usia`: INT
- `kelamin`: CHAR(10)
- `email`: VARCHAR(50)
- `PK_participants`: INT

## tests
- `id`: INT
- `participant_id`: INT
- `kb_id`: INT
- `literal`: VARCHAR(50)
- `value`: BOOL
- `PK_tests`: INT
```

Figure 3. Sequence Diagram

Figure 4. Database Design

3.3.4. Class Diagram Design
3.3.5. Forward Chaining Algorithm
Next is our forward chaining algorithm according to Russel and Norvig [12]:

function PL-FC-ENTAILS?(KB, q) returns true or false
    inputs: KB, the knowledge base, a set of propositional definite clauses
            q, the query, a proposition symbol
    count[c]rarr; a table, where count[c] is the number of symbols in
    c's premise
    inferred[rarr; a table, where inferred[s] is initially false for all symbols
agenda \&larr; a queue of symbols, initially symbols known to be true in KB

while agenda is not empty do

\hspace{1em} p \&larr; Pop(agenda)
\hspace{1em} if p = q then return true
\hspace{1em} if inferred[p] = false then
\hspace{2em} inferred[p] \&larr; true
\hspace{2em} for each clause c in KB where p is in c.PREMISE do
\hspace{3em} decrement count[c]
\hspace{3em} if count[c] = 0 then add c.CONCLUSION to agenda

return false

The implementation of the algorithm is using library that build from AIMA’s Project [12] as a jar (java archive) file, because we will use Java programming language. In order for this library to work in web application, then we will use Servlet and JSP technology and using tomcat as the server these are the simple and best technology according to Murach’s tutorial [13].

```java
public boolean plfcEntails(KnowledgeBase kb, PropositionSymbol q) {
  // count <- a table, where count[c] is the number of symbols in c's
  // premise
  Map<Clause, Integer> count = initializeCount(kb);
  // inferred <- a table, where inferred[s] is initially false
  for all // symbols
    Map<PropositionSymbol, Boolean> inferred = initializeInferred(kb);
  // agenda <- a queue of symbols, initially symbols known to be true in
  // KB
  Queue<PropositionSymbol> agenda = initializeAgenda(count);
  // Note: an index for p to the clauses where p appears in the
  premise
  Map<PropositionSymbol, Set<Clause>> pToClausesWithPInPremise =
    initializeIndex(
      count, inferred);

  // while agenda is not empty do
  while (!agenda.isEmpty()) {
    // p <- Pop(agenda)
    PropositionSymbol p = agenda.remove();
    // if p = q then return true
    if (p.equals(q)) {
      return true;
    }
    // if inferred[p] = false then
    if (inferred.get(p).equals(Boolean.FALSE)) {
      // inferred[p] <- true
      inferred.put(p, true);
      // for each clause c in KB where p is in c.PREMISE do
      for (Clause c : pToClausesWithPInPremise.get(p)) {
        // decrement count[c]
        decrement(count, c);
```
if count[c] == 0 {
    agenda.add(conclusion(c));
}

// return false
return false;

3.3.6. Interface Design

Figure 7. Registration Form

Figure 8. Symptoms Page
4. Conclusion
The result of the implementation works as we expected. The process starts from user fills up registration form and then continues to the symptoms page. User choose every symptom that appear in the page as they felt. It repeated until the last symptom. After that the forward chaining algorithm will start process the data from the user, and preview the result in summary page.

For further research we will update the knowledge base added from CDC symptoms. And we will suggest that the user must do PCR test for certain result, and compare the result with the expert system of Covid-19 detection.

4. Reference


