Fuzzy Based Decision Support Model for Health Insurance Claim

Sumiatie Susanto *[1], Ditdit Nugeraha Utama [1]
Email: sumiati.susanto@binus.ac.id, ditdit.utama@binus.edu
* Corresponding Author
[1] Computer Science Department, BINUS Graduate Program, Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480

Keywords: decision support model, decision support system, fuzzy logic, analytic hierarchy process, insurance claim

Received: August 6, 2022

Insurance industry in Indonesia has shown promising result based on premium growth in 2014-2018, as recorded in Indonesia General Insurance Market Update 2019. With the increase of premium, the claim rate also grows. Insurance companies face challenges in processing the claims. Many factors need to be carefully considered before making a claim decision. This paper proposes a decision support model (DSM) to score claim cases and to propose claim risk category (CRC) and claim decision (CD). The model was built with 13 parameters, divided into non-fuzzy group and fuzzy group. The analytic hierarchy process (AHP) method was used to determine the priority weight (PW) among parameters. The Tsukamoto's fuzzy logic (FL) method was applied to process the fuzzy parameters. A simple mathematics method (SMM) was exercised to calculate the non-fuzzy parameters, and to aggregate the result into claim risk score (CRS). Finally, CRC and CD were derived from the CRS using a rule base. The model was tested using 19611 actual claim history records. The result was: 6171 (31.47%) accepted with CRC= low, 3459 (17.64%) pending (CRC medium), and 9981 (50.89%) pending (CRC high). The DSM model was implemented in python with Google COLAB and Datapane to create various graphics.

Povzetek: Z metodo mehkih množic je narejen odločitveni sistem, ki preračuna tveganje in predlaga odobritev za zahtevek zdravstvenega zavarovanja.

1 Introduction

The development of the insurance industry in Indonesia over the past few years has shown promising improvement. Data from [1] revealed that during 2014-2018 the average premium growth was 15% and claim growth was 18% annually across the insurance industry. It also showed 10% and 9% respectively for premium and claim growth annually in general insurance (non-life) and re-insurance. According to [2], annual growth of Indonesia general insurance gross written premium was expected to reach 9.2% in 2026, after going through a sharp decline in 2020, impacted by the Covid-19 pandemic.

Along with the premium growth, the insurance claim rate also increases. One insurance company in Jakarta processed around 17,000 claims in 2020. Around 70% of these were claims for health insurance product with daily hospital reimbursement benefit, known as Hospital Cash Plan (HCP). The main challenge experienced by the company is how to produce CD with speed and accuracy. In 2020 the company achieved only 83% of its target claim processing time. This is due to the complexities of claim assessment process. There are many factors to be assessed to differentiate genuine claims from the potential fraudulent claims, before a claim assessor can make a right CD.

There are previous researches done to solve various areas in insurance industry. [3] showed many studies to solve problems in insurance such as underwriting classification, reserved funds for projected liabilities, reinsurance, pricing, asset and investment allocation using FL and variants of FL. The AHP and FL methods were used by [4] to create a model to determine the type of insurance product proposal suitable for potential buyers. A comparison of methods like AHP, technique for order of preference by similarity to ideal solution (TOPSIS) and simple additive weighting (SAW) was performed by [5] in a DSM case study to decide the eligibility of borrowers for financial institutions.

Specific studies in DSM to solve problems in insurance claim were also conducted by many researchers. A DSM based on AHP was created to determine the eligibility of surety bond insurance claims [6]. The genetic support vector machine approach was used to create a DSM to detect possible claim fraud [7]. A Bayesian quantile regression model made by [8] to detect which part of the claim distribution number has the greatest effect in vehicle insurance in Malaysia. A model to calculate claim reservation using fuzzy set theory was created by [9].

1.1 Related Works

Table 1 summarizes previous researches and the result. Apple-to-apple accuracy comparison might not be suitable because each model was created for a specific case and specific dataset. There is still a need for a model to support claim decision making for health insurance.

Previous researches resulted in a DSM based on certain methods suitable for each specific case and its dataset. This paper is to supplement researches in DSM, focusing on creating a model to suggest the right CD in health insurance claim. The novelty of this research is a method combining the AHP, the FL, and the SMM to create a multi-criteria rule-based DSM for claim decision. The contribution of this research is a model that is able to predict CD for the company. This is vital, because if a claim conclusion is wrong, it would give negative impact on customers and the business. Customers could be harmed by late or wrong verdict, and the corporation could suffer losses or reputational damage from wrong claim judgement.

This paper has 5 main sections. Section 1 is an introduction to the research. Section 2 discusses the material and methods in great detail. Section 3 displays the result and discussion. Section 4 is the conclusion and further work. Section 5 is a reference list cited in this paper.

Table 1 Previous Researches

Reference and Research Topic	Methods	Research Result
[3] FL and its variants used to solve many areas in insurance	FL and its variants	No stated accuracy result. FL gives more flexibility
[4] AHP and FL to create a model to determine insurance product proposal	AHP and FL	No stated accuracy result
[5] DSM case study comparing AHP, TOPSIS and SAW to decide the eligibility of borrowers	AHP, TOPSIS, SAW	AHP was said to produce better result in Euclidean distance analysis
[6] DSM to determine the eligibility of surety bond insurance claims	AHP	No stated accuracy result
[7] Compare 3 GSVM classifiers to create a DSM to detect possible claim fraud	GSVM	Linear (80.67%), Polynomial (81.22%), Radial Basis Function (87.91%)
[8] Compare Bayesian quantile, Poisson, and negative binomial regression to create a model to detect which part of the claim distribution number has the greatest effect in vehicle insurance in Malaysia	Bayesian, Poisson, and negative binomial regression	Bayesian overestimates the actual data by 0.79%, Poisson underestimates by 0.69%, and negative binomial overestimates by 3.65%
[9] Model to calculate claim reservation	Fuzzy Set Theory	No stated accuracy result

2 Material and Methods

As stated in the introduction, it is important for an insurance company to be able to correctly assess claim cases and issue a valid CD. A claim assessor must be able to identify potential frauds. According to [10] insurance fraud is an act that violates the law with the aim of getting financial benefits from an insurance company. There are multiple factors to be considered before accepting or rejecting a claim. Among them are: administrative completeness, suitability factor for medical services, accuracy in diagnosis, accuracy in disease codification according to international classification of diseases (ICD) [11].

DSM was chosen as the subject of this research to propose CD. DSM or modeling can help human make decisions that are logical, rational, structured and objective [12]. A model is a replica or imitation of a fact or a reality, it is not an actual fact or reality [13]. The purpose of a model is to explain something so that it is easier to be understood. Model development must be academically logical, meaning that model development must use methods that are valid and based on previously existing theories. Model must be factual, so that they can be analyzed, calculated, and producing predictions that can be verified and validated [13].

2.1 Research Methodology

The research methodology shown in Figure 1 was adopted from the seven stages of the DSM Wheel [12]. Problem or case analysis was carried out by conducting literature study & field study on DSM. Literature study was done on DSM techniques and how DSM could be used to solve insurance problems. Field study was conducted by studying the real case in the company. From this case analysis, the research goal was determined. The goal was to make the right and suitable DSM model to produce insurance claim decisions.

Next step was to analyze the decisions that will be proposed by the model. The model was to propose a CD, whether to accept or pending the claim. Pending means need further investigation by the claim assessor. The proposed CD was assessed depending on the potential risk of the claim, which was calculated as a CRS. CRS was categorized into a CRC of high, medium, or low risk. If CRC is low, the model will propose a CD to accept the claim. If the risk is medium or high, the model will propose a CD to pending the claim.

Parameter analysis is the process of analyzing what factors or criteria were used in the field for claim assessment, and what criteria was more important than others. This was done mainly by interviewing experts and literature study. There were 13 parameters in Table 2, defined by a team of 4 claim assessors whose experience was more than 5 years. Expert interview method has long been accepted in qualitative research, and is an efficient method [14]. There are several techniques to conduct interviews, such as face-to-face, telephone and text based

[15]. Given the pandemic Covid-19 situation, the interviews were conducted virtually, using Cisco WebEx platform. It is a collaboration platform where multiple participants can collaborate virtually, giving virtually similar experience as a face-to-face interview [16].

Data collection was done by obtaining historical claim data from the company. Then performing data cleansing, transforming, and formatting so it can be used as input for the model. Data cleansing was to remove some rows due to anomaly or incomplete data. For example, some columns were blank, or certain column values were not valid. Data transforming was to convert the non-fuzzy parameter value from non-numeric to numeric according to Table 3. For example, column gender has value "M" or "F" was converted to numeric 1 or 0 respectively. For fuzzy parameter in Table 4, some columns like Claim Amount and Daily Benefit were transformed to value in IDR thousand. Other columns with high / medium / low value were converted to 1, 0.5, and 0 respectively. Data formatting was to rename the columns according to the designed input file for the model. Final sample cleansed data can be seen in Table 6.

The next stage was to create the model. The model was built with 13 selected parameters, divided into non-fuzzy group (ax) and fuzzy group (bx). The AHP method was used to determine the PW among parameters. The Tsukamoto's FL method was applied to process the fuzzy parameters. An SMM was exercised to calculate the non-fuzzy parameters, and to aggregate the result into CRS. These methodologies were explained in the Framework Theory sub section.

Finally, the CD was derived from the CRS using a rule base. CRS was grouped into CRC low, medium, high. The model proposed to accept the claim when the CRC is low and to pending the claim for further investigation when the CRC is medium or high.

The model was validated and verified, and was tested using claim history data. The DSM model was implemented in python with Google COLAB platform and Datapane platform to create the various graphics.

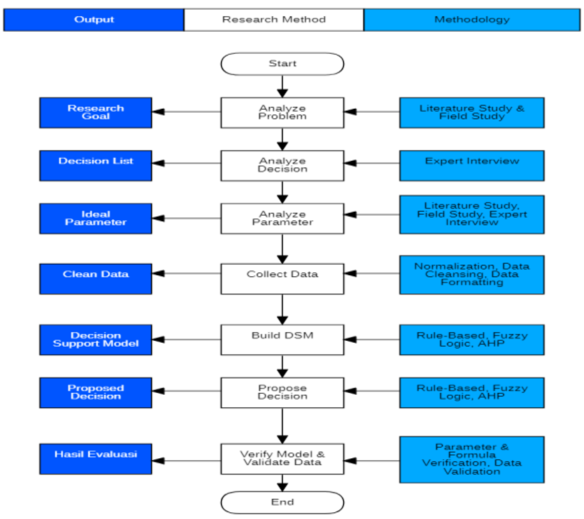

Figure 1: Research Methodology

2.2 Theory Framework

To create the model, unified modeling language (UML) was used to describe the model. UML is the industry standard language to describe and to visualize a model. It is commonly used in constructing and documenting an object-based system [17]. Some UML diagrams such as class diagram, activity diagram and influence diagram were selected to visualize the model. Sample diagrams can be seen in Figure 6-7 in Section 3.

AHP method [18] was used to determine the PW of the parameters. AHP is useful when there are many factors or criteria to be considered to make the right decision. According to [13], the AHP concept emphasizes the comparison of each criterion with every other criterion in terms of its level of importance. It was done by performing a pairwise comparison using a numerical rating or a value scale of 1-9 as shown in Table 5 [18].

It is necessary to check consistency ratio (CR) to ensure the PW comparison is consistent with each other. CR ≤ 0.1 means it is consistent, otherwise it is not, and the process must be repeated until it reaches consistency. Equation (1) and Equation (2) show the formula to calculate CR. CR is consistency ratio, CI is consistency index, RI is random index. n is the number of parameters used. Lambda max (λ max) was obtained first by performing matrix multiplication between the original pairwise comparison in Table 7 with the matrix PW in Table 8, resulting a new matrix Result (R) in Table 10. Then divide the value of R in each row with the PW of each row. Finally, the average value of λ max was taken as the final result of λ max [13]. The result can be seen in Table 10.

$$CR = CI \;/\; RI \qquad (1)$$
$$CI = (\lambda \max - n) \;/\; (n - 1) \qquad (2)$$

Non-fuzzy parameters were processed using AHP method and an SMM. Mathematical model is a description of a system using mathematical concepts and language [19]. Mathematical modeling is the process of building a model to explain a concept in a mathematical form so that it can be analyzed by performing mathematical calculations. According to [20] mathematical modeling includes the transition from a real world problem to a model representing it, then using that model to study and then solve the problem. Non-fuzzy parameters were assigned a numerical value as shown in Table 3. Then an SMM calculation was performed to obtain final value of non-fuzzy parameter (NF). It was done by multiplying the numeric value (ax) with the PW of each parameter (PW(ax)), then totaling them up.

Tsukamoto FL method was used to process the fuzzy parameters. FL is a logical concept to convert judgments in human language into a definite value (crisp). FL has been used in many domains such as in business, engineering, science, medical, and others. It was widely used [21] [22] [23], because its approach was more natural. It uses human language and imitates the concept of human thinking logic by using if-then rule-based in the decision-making process. FL was more tolerant to biased or uncertain data elements, as often found in reality. FL could model a complex uncertainty problem into a mathematical model for problem solving.

The FL approach uses fuzzy variables to represent linguistic expressions used by humans. Fuzzy variables are defined in the membership function (MF), describing the degree of membership of the variable in the fuzzy set. Three commonly used MF [22] are: linear up or linear down function, triangle function, and trapezoid function. MF in linear up or linear down is described in the form of a straight line that goes up or down. Triangle function is a combination of up and down linear function. Trapezoid is similar to triangle, with a horizontal top. Figure 2 shows an ascending linear curve with a lower bound a and an upper bound b. The exact input value of x can be less than a, or between a and b, or greater than b. The degree of membership x is represented by the symbol μ(x). To calculate μ(x) on an ascending linear curve, can be seen in Equation (3). Figure 3 shows a descending linear curve. Equation (4) shows how to calculate μ(x) on a descending linear curve. These 2 equations can be used to calculate μ(x) on triangle and trapezoid function. Figure 4-5 are the triangle and trapezoid function.

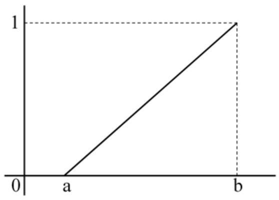

Figure 2: MF Linear Up

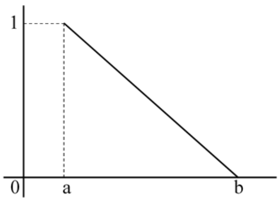

Figure 3: MF Linear Down

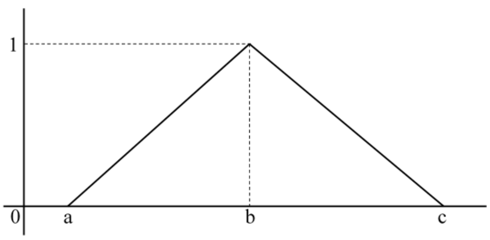

Figure 4: MF Triangle

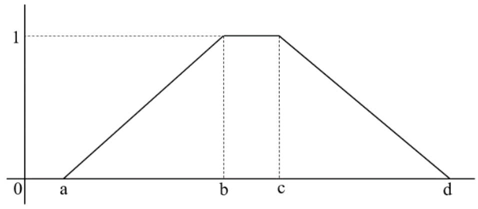

Figure 5: MF Trapezoid

$$\mu(x) = \begin{cases} 0, & x < \mathrm{a} \\ \dfrac{x - a}{\mathrm{b} - \mathrm{a}}, & \mathrm{a} \leq x < \mathrm{b} \\ 1, & x \geq \mathrm{b} \end{cases} \tag{3}$$

$$\mu(x) = \begin{cases} 1, & x < \mathrm{a} \\ \dfrac{\mathrm{b} - x}{\mathrm{b} - \mathrm{a}}, & \mathrm{a} \leq x < \mathrm{b} \\ 0, & x \geq \mathrm{b} \end{cases} \tag{4}$$

In addition to the membership function, it is necessary to know the operations on fuzzy sets for the inference process. The most commonly used are union and intersection operations. The union is an OR operation, takes the maximum value of (x) between the two sets. The intersection is an AND operation, takes the minimum value of (x) [22].

The stages of the FL algorithm [12], [22] can be seen in Figure 6. The initial input value is a definite value (crisp input). Then fuzzification is carried out to convert this definite value into a fuzzy input value. Followed by the inference process based on the rules, to get the fuzzy output value. Finally, defuzzification process is carried out to change the fuzzy output into crisp output.

There are several well-known FL methods such as the Tsukamoto method, the Sugeno method and the Mamdani method [22]. All three follow the steps of the FL process in Figure 6. The only difference is in the inference and defuzzification processes. The Tsukamoto method uses a firm If-Then rule and for defuzzification using the weighted average method. Equation (5) shows the Tsukamoto's inference rule and Equation (6) shows the Tsukamoto's defuzzification formula.

$$If\ (x\ is\ A)\ and\ (y\ is\ B)\ Then\ Z\ is\ C \tag{5}$$

$$Z = \frac{\sum(\alpha i\ X\ Zi)}{\sum(\alpha i)} \tag{6}$$

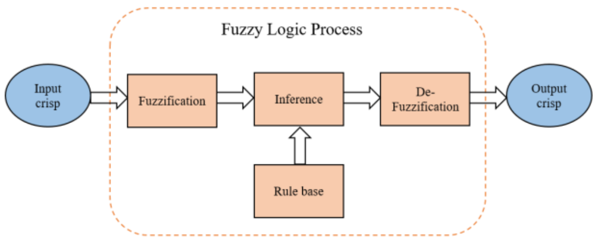

Figure 6: Fuzzy Logic Process

3 Result and Discussion

3.1 High Level Configuration of Model

Model entity diagram in Figure 7 shows the entities involved in the model. One Policy Holder has one or more health Insurance Policy. The Policy Holder is usually also the Insured, but not necessarily vice versa. An Insurance Policy may have zero or more Claim Transaction and zero or more Claim History. One Claim Transaction will have a Claim Decision consisting of a Claim_RiskScore and a Claim_Decision. Claim Transaction was calculated based on the AHP Method, Fuzzy Logic Method, and Simple Math Method. Fuzzy Logic Method consists of Language Variable that form a Fuzzy Membership function and a Rule Base.

Figure 8 is the algorithm diagram to explain the process flow of the model. First, define and analyze the parameters to be used. Then extract the claim transaction data. The parameters were then compared in pairs with the AHP method to obtain the PW of each parameter. Then the parameters were grouped into 2 groups based on data type and value. Non-fuzzy parameters (ax) are those that would not have bias value, and fuzzy parameters (bx) are those that could have bias value. Group (ax) were calculated using a SMM and AHP method, shown in Equation (7). Group (bx) were processed following the Tsukamoto Fuzzy Inference System method and multiplied by the total PW of (bx), shown in Equation (8). The CRS was an aggregation of the two groups, displayed in Equation (9). Then the CRS was evaluated according to a rule-based logic to determine a CD.

$$(NF) = \sum\big((ax) X\, PW(ax)\big) \quad (7)$$
$$(F) = Z(bx) X \sum\big(PW(bx)\big) \quad (8)$$
$$CRS = (NF) + (F) \quad (9)$$

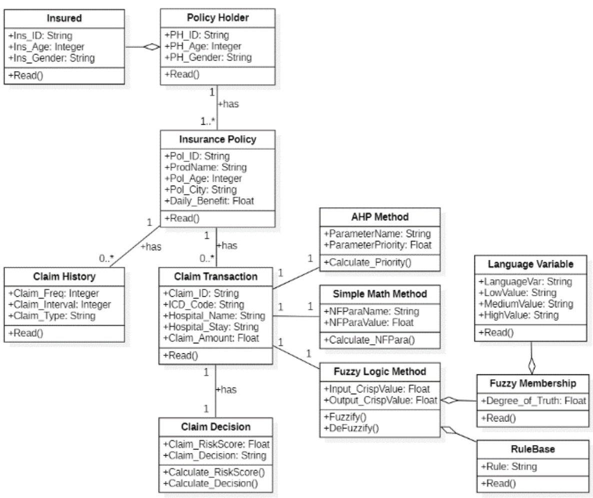

Figure 7: Model Entity Diagram

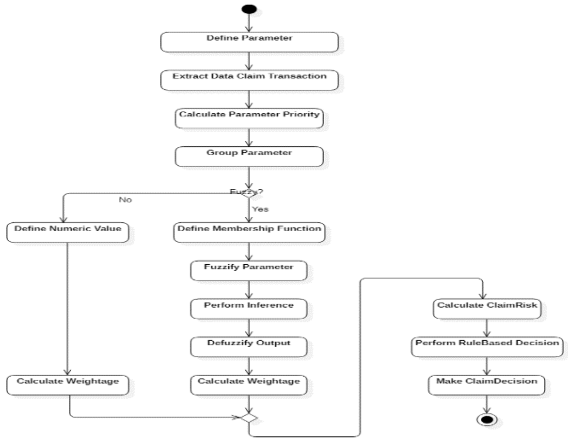

Figure 8: Model Algorithm Diagram

3.2 Parameter Decision

Table 2 shows the 13 parameters defined by the experts. Non-fuzzy parameters, shown in Table 3 with a numeric value assigned to each parameter. Fuzzy parameters, displayed in Table 4 with value range commonly found in real claim cases. A fuzzy membership range was created for each fuzzy parameter.

3.3 Collected Data

The result of data cleansing was shown in Table 6, with some sample data. For better visibility in this paper, the input column names were simplified to K1-K13 and the output column names to O1-O5. The actual column names were saved in the .CSV file format for the model.

Table 2: Parameter Table

Code	Criteria	Description
K1	Customer Age	Customer Age
K2	Customer Gender	Customer Gender
K3	City	City of policy issued
K4	Product Code	Insurance product code
K5	Policy Tenure	Months since policy was issued
K6	Claim Type	Cashless/reimbursement
K7	Claim Interval	Months since the last claim
K8	Claim Frequency	Claim frequency during policy tenure
K9	Days Hospitalized	Days of hospitalization
K10	Claim Amount	Claim amount in thousand IDR
K11	ICD-10	International Classification of Disease, 10th edition
K12	Daily Benefit	Daily cash benefit in thousand IDR
K13	Hospital Category	Hospital category (based on expert's opinion)

Table 3: Non-Fuzzy Parameter

Code	Value	Numeric Value
K2	Female Male	0 1
K3	City Risk Category: Low Medium High	0 0.5 1
K4	Product Risk Category: Low Medium High	0 0.5 1
K6	Cashless Reimbursement	0 0.5
K11	ICD-10 Category: Low Medium High	0 0.5 1
K13	Hospital Risk Category: Low Medium High	0 0.5 1

Table 4: Fuzzy Parameter

Code	Language Value	Reference Used in Actual Assessment	Fuzzy Member-ship
K1	Young Middle Mature	<= 32 years 33 – 44 >= 45	<= 35 28 – 48 >= 40
K5	Very New New Medium Long	0 – 4 month 5 – 13 14 – 28 >= 29	<= 6 4 – 14 10 – 30 >= 25
K7	Short Long	0 – 4 month > 4	<= 5 >= 3
K8	Seldom Often Very Often	<= 2 3 – 6 >= 7	<= 3 1 – 7 >= 5
K9	Short Medium Long	<= 3 days 4 >= 5	<= 4 2 – 6 >= 4
K10	Small Medium Large	< 3250 thousand 3250 – 6000 > 6000	<= 3500 2000–7000 >= 5500
K12	Low Medium High	<= 900 thousand 900 – 1000 > 1000	<= 950 700–1200 >= 950

3.4 Priority Weight

Pairwise comparison of the 13 parameters was carried out by the 4 experts, collaboratively producing an AHP matrix in Table 7. The experts filled the yellow cells, by rating the parameter importance in the row compared to the column. Example: row 1 of K1 (Customer Age) was compared to column 2 of K2 (Customer Gender), and was rated 5, meaning K1 was essentially more important than K2. In contrast, row 2 of K2 compared to column 1 of K1 was 1/5 = 0.2. This means K2 was essentially less important than K1. The green cells, were all 1, because they were a comparison of same parameter pairs. The bottom row was added to get the total value per column.

Table 7 was then normalized by dividing each value in Table 7 by the total value per column. Example: first cell in row K1 column K1, divided by total value of column K1 was 1 / 76.2 = 0.013. This process produced normalized value, recorded in Table 8. Total of each column was 1, meaning the values were proportionally correct. Then 1 column was added, to capture the average value of each row. This was the PW of each parameter in the row [13]. A new Table 9 was created, separating the group (ax) and (bx), sorted descending by PW.

CR calculation was done, following Equation (1) and (2). Table 10 shows the calculation of λ max, with result = 13.799. CI result = (13.779 – 13) / (13 – 1) = 0.065. RI was taken from the random index in Table 11 created by [18]. n is the number of parameter. For n=13, RI = 1.56. CR result = 0.065 / 1.56 = 0.042. It was $\leq$ 0.1, thus concluded that the pairwise comparison was consistent.

Table 5: Comparison Table [18]

Value	Description
1	Horizontal criteria is equally important as vertical criteria
3	Horizontal criteria is moderately more important than vertical criteria
5	Horizontal criteria is essentially or strongly more important than vertical criteria
7	Horizontal criteria is very strongly more important than vertical criteria
9	Horizontal criteria is extremely more important than vertical criteria
1/3	Horizontal criteria is moderately less important than vertical criteria
1/5	Horizontal criteria is essentially or strongly less important than vertical criteria
1/7	Horizontal criteria is very strongly less important than vertical criteria
1/9	Horizontal criteria is extremely less important than vertical criteria

Table 6: Claim Transaction

ClaimID	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	O1	O2	O3	O4	O5
2019-00001	41	1	1	1	18	0.5	18	1	2	8000	0	10000	0					
2019-00002	40	1	1	1	9	0.5	9	1	3	3000	1	1000	0					
…	…	…	…	…	…	…	…	…	…	…	…	…	…					
…	…	…	…	…	…	…	…	…	…	…	…	…	…					
2020-19610	30	1	1	1	26	0.5	4	6	5	6000	1	1000	1					
2020-19611	53	0	1	1	26	0.5	26	1	3	4000	0	5000	0					
2019-00001	41	1	1	1	18	0.5	18	1	2	8000	0	10000	0					

Table 7: Pairwise Comparison

Criteria	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13
K1	1,000	5,000	0,143	0,111	0,143	0,200	0,143	0,111	0,143	0,200	0,111	0,200	0,200
K2	0,200	1,000	0,111	0,111	0,111	0,143	0,111	0,111	0,111	0,143	0,111	0,143	0,143
K3	7,000	9,000	1,000	0,333	1,000	3,000	1,000	0,333	1,000	3,000	0,333	3,000	3,000
K4	9,000	9,000	3,000	1,000	3,000	5,000	3,000	1,000	3,000	5,000	1,000	5,000	5,000
K5	7,000	9,000	1,000	0,333	1,000	3,000	1,000	0,333	1,000	3,000	0,333	3,000	3,000
K6	5,000	7,000	0,333	0,200	0,333	1,000	0,333	0,200	0,333	1,000	0,200	1,000	1,000
K7	7,000	9,000	1,000	0,333	1,000	3,000	1,000	0,333	1,000	3,000	0,333	3,000	3,000
K8	9,000	9,000	3,000	1,000	3,000	5,000	3,000	1,000	3,000	5,000	1,000	5,000	5,000
K9	7,000	9,000	1,000	0,333	1,000	3,000	1,000	0,333	1,000	3,000	0,333	3,000	3,000
K10	5,000	7,000	0,333	0,200	0,333	1,000	0,333	0,200	0,333	1,000	0,200	1,000	1,000
K11	9,000	9,000	3,000	1,000	3,000	5,000	3,000	1,000	3,000	5,000	1,000	5,000	5,000
K12	5,000	7,000	0,333	0,200	0,333	1,000	0,333	0,200	0,333	1,000	0,200	1,000	1,000
K13	5,000	7,000	0,333	0,200	0,333	1,000	0,333	0,200	0,333	1,000	0,200	1,000	1,000
Total	**76,200**	**97,000**	**14,587**	**5,356**	**14,587**	**31,343**	**14,587**	**5,356**	**14,587**	**31,343**	**5,356**	**31,343**	**31,343**

Table 8: Normalized Value

Criteria	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	PW
K1	0,013	0,052	0,010	0,021	0,010	0,006	0,010	0,021	0,010	0,006	0,021	0,006	0,006	**0,015**
K2	0,003	0,010	0,008	0,021	0,008	0,005	0,008	0,021	0,008	0,005	0,021	0,005	0,005	**0,010**
K3	0,092	0,093	0,069	0,062	0,069	0,096	0,069	0,062	0,069	0,096	0,062	0,096	0,096	**0,079**
K4	0,118	0,093	0,206	0,187	0,206	0,160	0,206	0,187	0,206	0,160	0,187	0,160	0,160	**0,172**
K5	0,092	0,093	0,069	0,062	0,069	0,096	0,069	0,062	0,069	0,096	0,062	0,096	0,096	**0,079**
K6	0,066	0,072	0,023	0,037	0,023	0,032	0,023	0,037	0,023	0,032	0,037	0,032	0,032	**0,036**
K7	0,092	0,093	0,069	0,062	0,069	0,096	0,069	0,062	0,069	0,096	0,062	0,096	0,096	**0,079**
K8	0,118	0,093	0,206	0,187	0,206	0,160	0,206	0,187	0,206	0,160	0,187	0,160	0,160	**0,172**
K9	0,092	0,093	0,069	0,062	0,069	0,096	0,069	0,062	0,069	0,096	0,062	0,096	0,096	**0,079**
K10	0,066	0,072	0,023	0,037	0,023	0,032	0,023	0,037	0,023	0,032	0,037	0,032	0,032	**0,036**
K11	0,118	0,093	0,206	0,187	0,206	0,160	0,206	0,187	0,206	0,160	0,187	0,160	0,160	**0,172**
K12	0,066	0,072	0,023	0,037	0,023	0,032	0,023	0,037	0,023	0,032	0,037	0,032	0,032	**0,036**
K13	0,066	0,072	0,023	0,037	0,023	0,032	0,023	0,037	0,023	0,032	0,037	0,032	0,032	**0,036**
Total	**1,000**	**1,000**	**1,000**	**1,000**	**1,000**	**1,000**	**1,000**	**1,000**	**1,000**	**1,000**	**1,000**	**1,000**	**1,000**	**1,000**

Table 9: Priority Weight

Non-Fuzzy Group (ax)	Fuzzy Group (bx)	Non-Fuzzy Group (ax)	Fuzzy Group (bx)
Criteria Code	**Priority Weight**	**Criteria Code**	**Priority Weight**
K4	0,172	K8	0,172
K11	0,172	K5	0,079
K3	0,079	K7	0,079
K6	0,036	K9	0,079
K13	0,036	K10	0,036
K2	0,010	K12	0,036
		K1	0,015
Total (∑(PW(ax))	**0,504**	**Total (∑(PW(bx))**	**0,496**

Table 10: λ Max Calculation

Comparison matrix in Table 7					PW in Table 8	Result (R)	λ max
Criteria	**K1**	**…**	**K13**				
K1	1,000	…	0,200		0,015	0,194	13,140
K2	0,200	1,000	0,143		0,010	0,125	13,168
K3	7,000	…	3,000		0,079	1,110	14,029
K4	9,000	…	5,000		0,172	2,404	14,003
K5	7,000	…	3,000	**X**	0,079	1,110	14,029
K6	5,000	…	1,000		0,036	0,493	13,674
K7	7,000	…	3,000		0,079	1,110	14,029
K8	9,000	…	5,000		0,172	2,404	14,003
K9	7,000	…	3,000		0,079	1,110	14,029
K10	5,000	…	1,000		0,036	0,493	13,674
K11	9,000	…	5,000		0,172	2,404	14,003
K12	5,000	…	1,000		0,036	0,493	13,674

K13	5,000	...	1,000		0,036	0,493	13,674
Total	**76,200**	**...**	**31,343**		**1,000**	**Average λ max =**	**13,779**

Table 11: Random Index [18]

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0,00	0,00	0,58	0,90	1,12	1,24	1,32	1,41	1,45	1,49	1,51	1,48	1,56	1,57	1,59

3.5 Fuzzification, Inference, and Defuzzification

For each fuzzy parameter in this study, the fuzzy MF was determined first, as shown in Table 4. The grouping of each MF was based on field data and claim assessors' experience. Sample membership function of some parameters such as K5, K7 and K8 and the output Claim Risk membership can be seen in Figure 9-12. Equation (10) – (21) show the formula to calculate the (µx) in fuzzification process.

Next, inference process was performed following Equation (5). A rule-based was defined by the 4 experts. Initially there were 1944 rules, with many similarities. Then they were summarized into 13 final rules. Some samples of the rules were recorded in Table 12.

Last step was defuzzification process, to get the Claim Risk (Z(bx)) value using Equation (6). Then Z(bx) was multiplied by the total priority weight (∑(PW(bx))) as stated in Equation (8) to get the final fuzzy parameter result (F). The CRS was obtained from the sum of the non-fuzzy parameter values (NF) and fuzzy parameters (F) as shown in Equation (9). The CRS was then evaluated against a CD rule-based logic in Table 13 to get the CRC and CD.

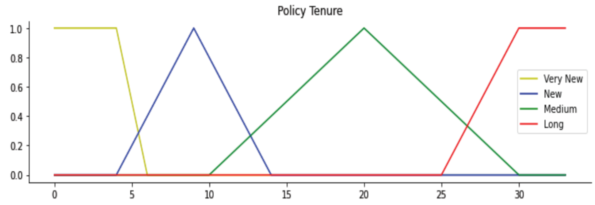

Figure 9: MF Policy Tenure (K5)

$$\mu VeryNew(x) = \begin{cases} 1, & x < 4 \\ \dfrac{6-x}{6-4}, & 4 \le x < 6 \\ 0, & x \ge 6 \end{cases} \tag{10}$$

$$\mu New(x) = \begin{cases} 0, & x < 4 \text{ or } x \ge 14 \\ \dfrac{x-4}{9-4}, & 4 \le x < 9 \\ 1, & x = 9 \\ \dfrac{14-x}{14-9}, & 9 < x < 14 \end{cases} \tag{11}$$

$$\mu Medium(x) = \begin{cases} 0, & x < 10 \text{ or } x \ge 30 \\ \dfrac{x-10}{20-10}, & 10 \le x < 20 \\ 1, & x = 20 \\ \dfrac{30-x}{30-20}, & 20 < x < 30 \end{cases} \tag{12}$$

$$\mu Long(x) = \begin{cases} 0, & x < 25 \\ \dfrac{x-25}{30-25}, & 25 \le x < 30 \\ 1, & x \ge 30 \end{cases} \tag{13}$$

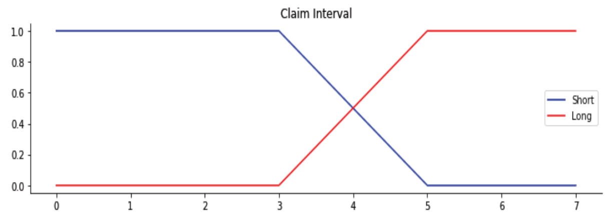

Figure 10: MF Claim Interval (K7)

$$\mu Short(x) = \begin{cases} 1, & x < 3 \\ \dfrac{5-x}{5-3}, & 3 \le x < 5 \\ 0, & x \ge 5 \end{cases} \tag{14}$$

$$\mu Long(x) = \begin{cases} 0, & x < 3 \\ \dfrac{x-3}{5-3}, & 3 \le x < 5 \\ 1, & x \ge 5 \end{cases} \tag{15}$$

Figure 11: MF Claim Frequency (K8)

$$\mu Seldom(x) = \begin{cases} 1, & x < 1 \\ \dfrac{3-x}{3-1}, & 1 \le x < 3 \\ 0, & x \ge 3 \end{cases} \tag{16}$$

$$\mu Often(x) = \begin{cases} 0, & x < 1 \text{ or } x \ge 7 \\ \dfrac{x-1}{4-1}, & 1 \le x < 4 \\ 1, & x = 4 \\ \dfrac{7-x}{7-4}, & 4 < x < 7 \end{cases} \tag{17}$$

$$\mu VeryOften(x) = \begin{cases} 0, & x < 5 \\ \dfrac{x-5}{7-5}, & 5 \le x < 7 \\ 1, & x \ge 7 \end{cases} \tag{18}$$

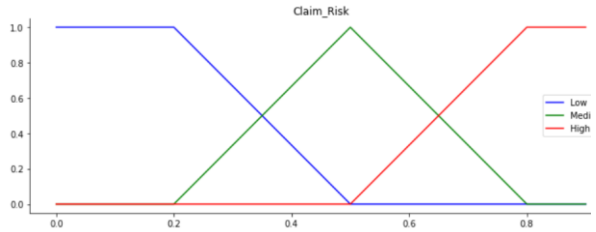

Figure 12: MF Claim Risk

$$\mu Low(x) = \begin{cases} 1, & x < 0{,}2 \\ \dfrac{0{,}5 - x}{0{,}5 - 0{,}2}, & 0{,}2 \le x < 0{,}5 \\ 0, & x \ge 0{,}5 \end{cases} \quad (19)$$

$$\mu Medium(x) = \begin{cases} 0, & x < 0{,}2 \text{ or } x \ge 0{,}8 \\ \dfrac{x - 0{,}2}{0{,}5 - 0{,}2}, & 0{,}2 \le x < 0{,}5 \\ 1, & x = 0{,}5 \\ \dfrac{0{,}8 - x}{0{,}8 - 0{,}5}, & 0{,}5 < x < 0{,}8 \end{cases} \quad (20)$$

$$\mu High(x) = \begin{cases} 0, & x < 0{,}5 \\ \dfrac{x - 0{,}5}{0{,}8 - 0{,}5}, & 0{,}5 \le x < 0{,}8 \\ 1, & x \ge 0{,}8 \end{cases} \quad (21)$$

Table 12: Final Inference Rule Based

RULE NO	IF	THEN CR =
1	K8= SELDOM AND K5 = VERY NEW AND K7= SHORT AND (K9 = SHORT OR K9 = MEDIUM OR K9 = LONG) AND (K10 =SMALL OR K10 = MEDIUM OR K10 = LARGE) AND (K12 = LOW OR K12 = MEDIUM OR K12 = HIGH) AND (K1 = YOUNG OR K1 = MIDDLE OR K1 = MATURE)	High
…	…	…
6	K8 = SELDOM AND K5 = MEDIUM AND (K7 = SHORT OR K7 = LONG) AND (K9 = SHORT OR K9 = MEDIUM) AND (K10 = SMALL OR K10 = MEDIUM OR K10 = LARGE) AND (K12 = LOW OR K12 = MEDIUM OR K12 =HIGH) AND (K1 = YOUNG OR K1 = MIDDLE OR K1 = MATURE)	Low
…	…	…
13	K8 = VERY OFTEN AND K5 = LONG AND K7 = LONG AND (K9 = SHORT OR K9 = MEDIUM) AND (K10 = SMALL OR K10 = MEDIUM OR K10 = LARGE) AND (K12 = LOW OR K12 = MEDIUM OR K12 = HIGH) AND (K1 = YOUNG OR K1 = MIDDLE OR K1 = MATURE)	Medium

Table 13: Claim Decision Rule

Rule ID	If Claim Risk Score	Then CR Category	Then Claim Decision
1	< 0,600	Low	Accept
2	≥ 0.600 and < 0,650	Medium	Pending
3	≥ 0,650	High	Pending

Table 14: Calculation Result from Model

ClaimID	NonFuzzyValue	FuzzyValue	ClaimRiskScore	ClaimRiskCategory	ClaimDecision
2019-00001	0.278	0.144	0.422	LOW	ACCEPT
2019-00002	0.450	0.159	0.609	MEDIUM	PENDING
…	…	…	…	…	…
2020-19610	0.486	0.222	0.708	HIGH	PENDING
2020-19611	0.269	0.198	0.467	LOW	ACCEPT

3.6 Proposed Decision

The model was run with 19611 claim history records. It proposed 6171 records (31.47%) with CRC = low and CD = accepted, 3459 records (17.64%) with CRC = medium and CD = pending, and 9981 records (50.89%) with CRC = high and CD = pending. Calculation result from the model was shown in Table 14 with some sample rows. Graphical dashboard with different views were displayed in Figure 13-19.

3.7 Discussion

Compared to previous studies related to insurance, some were using FL only [3] [9] or AHP only [6]. Some were using other methods or fewer parameters [5] [7] [8] [11]. Another combining AHP and FL but with only 4 parameters. This study was combining AHP, FL and SMM with 6 non-fuzzy parameters and 7 fuzzy parameters which made it more comprehensive. Other study that combined the 3 methods was done by [24] to determine student's academic performance. However, the non-fuzzy group (ax) was calculated by multiplying total group(ax) with total PW(ax). This paper was done by multiplying individual value of each parameter (ax) with individual PW(ax), then summed it up as total NF. This was more proportional and accurate.

Accuracy check of the model result is displayed in Table 15. Result from model was compared to the actual claim history result by claim assessor. Model result was

90.73% true positive where CD from model = accept with CRC = low and actual claim result = accepted. 9.49% can be classified as true positive where CD from model = pending with CRC = high (logically expected to be rejected) and actual claim result = rejected. Note that actual claim result does not have pending decision because already final decision.

Table 15 Model Accuracy Check

Model Result		Actual Claim Decision		Total	True Positive
CD	CRC	Accept	Reject		
Accept	Low	5599	572	6171	90.73%
Pending	Medium	3211	248	3459	
	High	9034	947	9981	9.49%
Total		17844	1767	19611	

4 Conclusion and Further Works

This research concluded that the model was able to produce the CRC of low / medium / high and the final CD as expected. The CRC will help claim assessors in distributing the cases among the assessors. For example, the low / medium risk to junior assessors and the high risk to senior ones. For further research, it would be good to add machine learning to enhance the model logic, and to add / remove parameters according to the real situation evolved in the future.

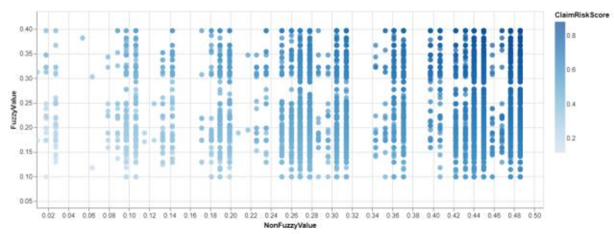

Figure 13: CRS Distribution Dashboard

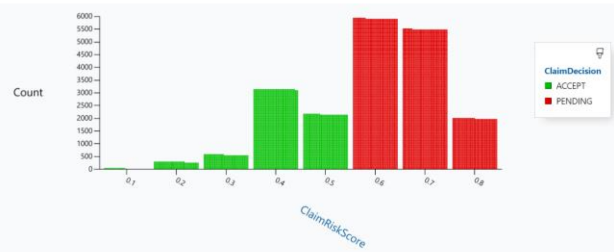

Figure 14: CD by CRS

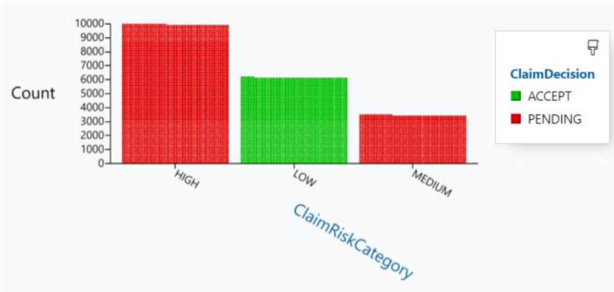

Figure 15: CD by CRC

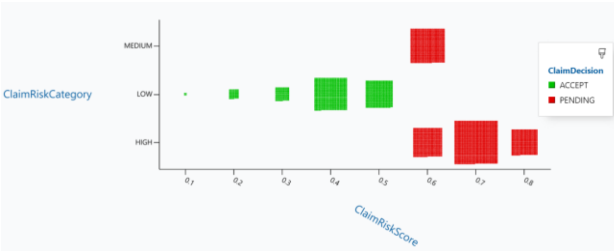

Figure 16: Density graph by CRS and CDC

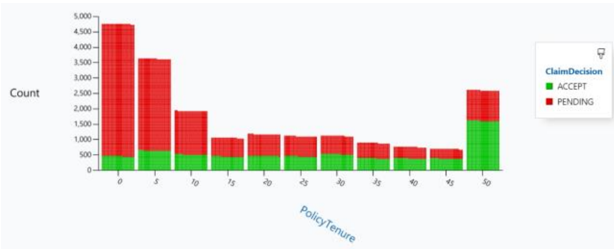

Figure 17: CD by Policy Tenure

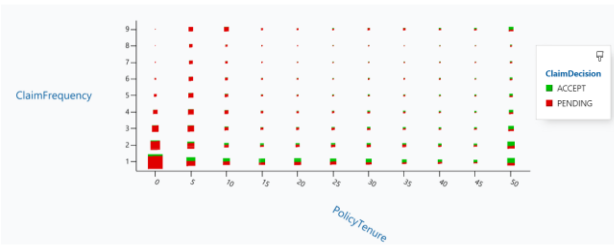

Figure 18: Density graph by Claim Frequency and Policy Tenure

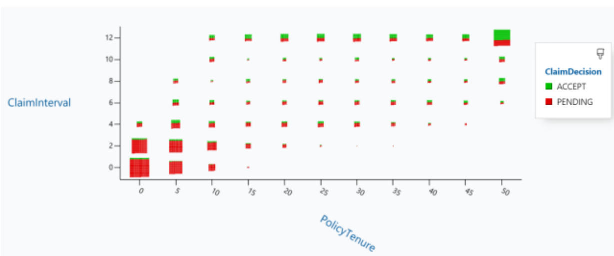

Figure 19: Density graph by Claim Interval and Policy Tenure

References

[1] Asosiasi Asuransi Umum Indonesia, "Indonesia General Insurance Market Update 2019," Asosiasi Asuransi Umum Indonesia, Jakarta, 2019.

[2] GlobalData, "Globaldata.com," 11 February 2022. [Online]. Available: https://www.globaldata.com/driven-economic-reforms-indonesian-general-insurance-market-reach-6-3bn-2025-finds-globaldata/. [Accessed 1 Apr 2022].

[3] A. F. Shapiro, "An Overview of Insurance Uses of Fuzzy Logic," *Computational Intelligence in Economics and Finance,* p. 25–61, 2007. https://doi.org/10.1007/978-3-540-72821-4_2

[4] C.-S. Huang, Y.-J. Lin and C.-C. Lin, "An Evaluation Model for Determining Insurance Policy Using AHP and Fuzzy Logic: Case Studies of Life and Annuity Insurances," in *8th WSEAS International Conference on Fuzzy Systems,* Vancouver, British Columbia, Canada, 2007.

[5] D. Himawan, "Analisis Perbandingan Menggunakan Metode AHP, TOPSIS, dan SAW dalam Studi Kasus Sistem Pendukung Keputusan Peminjam yang Layak Bagi Lembaga Keuangan," *repositori.usu.ac.id,* 2019.

[6] A. I. Islam, A. Jamaludin and N. Heryana, "Sistem Pendukung Keputusan Kelayakan Klaim Asuransi Menggunakan Metode AHP," *Jurnal Informatika Polinema,* vol. 7, no. 2, pp. 115-122., 2021. https://doi.org/10.33795/jip.v7i2.398

[7] R. A. Sowah, M. Kuuboore, A. Ofoli, S. Kwofie, L. Asiedu, K. M. Koumadi and K. O. Apeadu, "Decision Support System (DSS) for Fraud Detection in Health Insurance Claims Using Genetic Support Vector Machines (GSVMs)," *Journal of Engineering,* p. 1–19, 2019. https://doi.org/10.1155/2019/1432597

[8] M. F. M. Fuzi, A. A. Jemain and N. Ismail, "Bayesian quantile regression model for claim count data," *Insurance: Mathematics and Economics,* vol. 66, p. 124–137, 2016. https://doi.org/10.1016/j.insmatheco.2015.11.004

[9] J. d. A. Sánchez, "Calculating insurance claim reserves with fuzzy regression. Fuzzy Sets and Systems," *Science Direct,* vol. 157, no. 23, p. 3091–3108, 2006. https://doi.org/10.1016/j.fss.2006.07.003

[10] Y. Sarwo, "Tinjauan Yuridis Terhadap Kecurangan (Frauds) Dalam Industri Asuransi Kesehatan di Indonesia," *Kisi Hukum,* vol. 14(1) , pp. 78-92, 2015.

[11] Y. Yulia and H. Putri, "Literatur Riview Tentang Faktor Penyebab Klaim Tidak Layak Bayar BPJS Kesehatan Di Rumah Sakit Tahun 2020," *Jurnal Ilmiah Perekam dan Informasi Kesehatan Imelda (JIPIKI),* vol. 6(1), pp. 83-90., 2021. https://doi.org/10.52943/jipiki.v6i1.487

[12] D. N. Utama, Logika Fuzzy untuk Model Penunjang Keputusan, Yogyakarta: Garudhawaca, 2021.

[13] D. N. Utama, Sistem Penunjang Keputusan: Filosofi, Teori dan Implementasi, Yogyakarta: Garudhawaca, 2017.

[14] A. Bogner, B. Littig and W. Menz, "Introduction: Expert interviews—An introduction to a new methodological debate. In Interviewing experts," *Palgrave Macmillan,* pp. 1-13, 2009. https://doi.org/10.1057/9780230244276_1

[15] R. Opdenakker, "Advantages and disadvantages of four interview techniques in qualitative research.," *Forum qualitative sozialforschung/forum: Qualitative social research,* vol. 7, no. 4, 2006. https://doi.org/10.17169/fqs-7.4.175

[16] R. Singh and S. Awasthi, "Updated comparative analysis on video conferencing platforms-zoom, Google meet, Microsoft Teams, WebEx Teams and GoToMeetings," *EasyChair Preprint,* vol. 4026, pp. 1-9, 2020.

[17] L. Mathiassen, A. Munk-Madsen, P. A. Nielsen and J. Stage, Object Oriented Analysis & Design, Aalborg: Marko Publishing ApS, 2000.

[18] R. Saaty, "The analytic hierarchy process—what it is and how it is used.," *Mathematical modelling,* Vols. 9(3-5), pp. 161-176, 1987. https://doi.org/10.1016/0270-0255(87)90473-8

[19] "Mathematical Model," 2021. [Online]. Available: https://en.wikipedia.org/wiki/Mathematical_model.

[20] R. Crouch and C. Haines, "Mathematical modelling: Transitions between the real world and the mathematical model," *International Journal of Mathematical Education in Science and Technology,* vol. 35 , no. 2, pp. 197-206, 2004. https://doi.org/10.1080/00207390310001638322

[21] Yulmaini, Logika Fuzzy Studi Kasus & Penyelesaian Menggunakan Microsoft Excel dan Matlab, Yogyakarta: Penerbit ANDI, 2018.

[22] E. Hidayat, "Logika Fuzzy [1]: Fungsi Keanggotaan," Jakarta, 2020. https://www.youtube.com/watch?v=orVFpa3fhB8

[23] L. A. Zadeh, "Soft computing and fuzzy logic.," *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A Zadeh,* pp. 796-804, 1996. https://doi.org/10.1109/52.329401

[24] D. Kurniawan and D. N. Utama, "Decision Support Model using FIM Sugeno for Assessing the Academic Performance," *Advances in Science, Technology and Engineering Systems Journal,* vol. 6, pp. 605-611, 2021. https://dx.doi.org/10.25046/aj060165