

# Prediction of the Number of New Cases of Covid-19 in Indonesia Using Fuzzy Time Series Model Chen

Kanda Januar Miraswan<sup>1</sup>, Wiwik Anum Puspita<sup>2,\*</sup>, Alvi Syahrini Utami<sup>3</sup>

Informatics Department, Computer Science Faculty, Universitas Sriwijaya

<sup>1</sup>kandajm@ilkom.unsri.ac.id; <sup>2</sup>wiwikiumpuspita@gmail.com\*; <sup>3</sup>alvisyahrini@ilkom.unsri.ac.id

## ARTICLE INFO

### Article history

Received 19 Nov 2021

Revised 27 Nov 2021

Accepted 3 Dec 2021

### Keywords

Number of new cases

Covid-19

Fuzzy Time Series Chen

MAPE

## ABSTRACT

Coronavirus Diseases 2019 (Covid-19) is a disease caused by a virus that originated in Wuhan, China. This virus infects people rapidly to the country of Indonesia. According to the latest Covid-19 Development Team in Indonesia, as of 09/08/2021, there were around 3,686,740 people who were confirmed positive for Covid-19. With the numbers continuing to grow, predictions of new cases of Covid-19 in Indonesia were made using the time series method. The method used by the researcher is Chen's Fuzzy Time Series. The purpose of the researcher is to forecast, to find out the prediction of the number of new cases of Covid-19 in Indonesia using the FTS Chen method into software. In addition, in order to provide information to predict, so that the government knows and can make decisions. To measure the performance of the method, the Mean Absolute Percentage Error (MAPE) is used as a measure of the level of accuracy of the forecasting performed. The test data used were 363 data with several variations of parameters D1 & D2. From the results of the analysis that was tested by the researcher, with 50 trials of parameter input, better accuracy results were obtained at input D1 = 135135 and D2 = 2000 with MAPE is 35.55006797 (35%).

## 1. Introduction

Coronavirus Disease 2019 or also known as Covid-19 is a virus that originated in Wuhan, China. This virus is spread by infecting humans to humans through inhalation. Covid-19 was first reported on December 31, 2019. This virus spread rapidly in China, within about two or three months Covid-19 spread in several countries including Asia, Europe, America, Australia and Africa. [1].

The arrival of Covid -19 in the country of Indonesia has made Indonesian citizens attacked by a dangerous virus. The number of patients in Indonesia is increasing day by day, until it extends to cities in Indonesia [2]. With the grow up and spread that is expanding rapidly, a prediction is needed. With predictions, it can be information that will be useful for decision makers in dealing with the spread of Covid-19 cases in Indonesia.

Prediction is a systematic estimate that information can be taken from data in the past. Prediction can also be called forecasting. Forecasting has two techniques, namely quantitative analysis and qualitative analysis. Quantitative analysis is an analysis obtained based on an explanation or statement from someone that cannot be converted into data in the form of numbers. While qualitative analysis is the analysis of data taken from the past, the data is in the form of numbers or called time series data [3]. So predicting the number of new Covid-19 cases is better using time series analysis, because analyzing using the time series can help estimate future values.

Fuzzy Time Series (FTS) is a well-known forecaster to predict using historical data. FTS has been implemented by many previous researchers, because it has a system calculation process that is not too complicated. One of them is Chen's Fuzzy Time Series model. Chen in 1996 used arithmetic operations. Fuzzy Time Series is many used from other time series methods, namely the values used during forecasting. Fuzzy Time Series forecasting has a fuzzy set of actual numbers over a set universe. The fuzzy set has the meaning as a class of numbers with vague boundaries. Fuzzy Time Series is also widely applied to Historical Data, because it can solve forecasting problems which are not only in the form of real numbers but can be applied to linguistic values. Some of the advantages of using FTS are: the prediction process is presented in the form of linguistic data, FTS is famous for its effective methods, the forecasting process is based on time series data analysis, and has a good level of accuracy [4].

The basis for taking using the Chen model is carried out by researchers using a variable that states the number of residents who are confirmed to be Covid-19 can be done with good value results. In this study, we can determine the Fuzzy Time Series Chen method in the case of predicting the number of Covid-19 cases in Indonesia.

## 2. Literature Study / Hypotheses Development

### a. Fuzzy Time Series model Chen

Steps in using Chen's Fuzzy Time Series Model:

#### 1. Determine Universe of Discourse

The first way before determining the Universe  $U$  is to determine the maximum value of ( $X_{max}$ ) and minimum ( $X_{min}$ ) obtained from historical data. After that, determine the two positive numbers determined by the researcher. According to the journal [5], then use the following formula:

$$U = [X_{min} - D_1, X_{max} + D_2] \quad (1)$$

- $U$  = Universe of discourse
- $X_{min}$  = Minimum value of historical data
- $D_1$  = Proper positive number 1
- $X_{max}$  = Maximum value of historical data
- $D_2$  = Proper positive number 2

According to the theoretical basis of the researchers who previously calculated using the Chen FTS method, there were 2 positive numbers, namely  $D_1$  and  $D_2$ . Input value  $D_1$  and  $D_2$  has uses that greatly affect the value of  $U_{max}$  and  $U_{min}$ , which is where different inputs will produce different forecast values. For  $D_1$  can widen the lower limit of the interval and  $D_2$  can widen the upper limit of internal [3].

Random value determination for  $D_1$  &  $D_2$  get references from journals [3] in the journal stated that research in the case of forecasting affects the determination of the *error* value. The researcher conducted an influence test with 4 distributions of range values for  $D_1$  small,  $D_1$  big,  $D_2$  small and  $D_2$  big. With conditions  $D_1$  small = 0-50,  $D_1$  big = 51-1000,  $D_2$  small = 0-100 and  $D_2$  big = 101-1000. Researchers conducted some of these tests found the smallest error value was in the range <10% with positive number input  $D_1$  big = 100 and  $D_2$  small = 0.

#### 2. Determine the Number of classes and Length of the Internal

How to determine the number of classes is by the formula [6] :

$$\text{Number of intervals} = 1 + 3,322 \log n \quad (2)$$

- $n$  = amount of actual data used

After the number of classes is determined, then determine the length of the interval using the formula [6] :

$$Interval\ length = \frac{D_{max}-D_{min}}{Interval} \tag{3}$$

$D_{max}$  = the largest actual data plus the second positive number 2

$D_{min}$  = the smallest actual data is subtracted by any positive number 1

After the number of classes and the length of the class are determined, then divide the group of data into the number of intervals by making values for the Lower Limit, Upper Limit, and Middle Value as much as the Class that is obtained. Due to simplify the next step.

### 3. Define Fuzzy Set

The fuzzy set is also known as the basis of fuzzy logic. A fuzzy set is a set that has linguistic variables and a set whose elements have degrees of membership. A linguistic variable is a variable that defines a condition that is expressed by a membership function in the universe U, where a membership value is expressed in degrees of membership. [7].

In order for the data to become a linguistic variable, a fuzzy set is carried out  $A_1, A_2, A_3, A_4, \dots, A_n$  as follows:

$$\begin{aligned} A_1 &= 1/u_1 + 0,5/u_2 + 0/u_3 + \dots + 0/u_n \\ A_2 &= 0,5/u_1 + 1/u_2 + 0,5/u_3 + \dots + 0/u_n \\ A_3 &= 0/u_1 + 0,5/u_2 + 1/u_3 + \dots + 0/u_n \\ A_i &= 0/u_1 + 0/u_2 + 0/u_3 + \dots + 0,5/u_{n-1} + 1/u_n \end{aligned} \tag{4}$$

Each every fuzzy set  $A_i$  made as many as the number of intervals that have been set in the previous stage [8] or can be described by the matrix below:

**Table 1.** Fuzzy Set Martix

$A_{ij}$	$A_{1i}$	$A_{2i}$	$A_{3i}$	$A_{4i}$	$A_{.i}$
$A_{1j}$	1	0,5	0	0	..
$A_{2j}$	0,5	1	0,5	0	..
$A_{3j}$	0	0,5	1	0,5	..
$A_{4j}$	0	0	0,5	1	..
$A_{.j}$	..	..	..	..	..

In Chen's Fuzzy Time Series model, there are several rules [8] :

- a. Rule 1 : If historical data is included in  $u_i$ , then the degree of membership is 1, and  $u_i + 1$  is 0,5 and if not both then stated 0.
- b. Rule 2 : If historical data is included in  $u_i$ , with a statement like this  $1 \leq i \leq p$  then the value of the degree of membership for  $u_i$  is 1, for  $u_i - 1$  and  $u_i + 1$  is 0,5 and if not all three then stated 0.
- c. Rule 3 : If historical data is included in  $u_p$ , then the degree of membership is 1, for  $u_p - 1$  is 0,5 and if not both then stated 0.

#### 4. Determine Fuzzy Logic Relations (FLR)

FLR is created as a relation performed based on historical data value  $A_i$ .  $FLR A_i \rightarrow A_j$  means to make groups according to time, where  $A_i$  are year  $i$  and  $A_j$  year  $i+1$ .  $A_i$  acting as the left side of the relationship is called the current state and  $A_j$  and acting as the right side of the relationship as the next state. When a relation is repeated, it is still counted once [5].

#### 5. Determine Fuzzy Logic Relations Group (FLRG)

At the FLRG stage, it is only to classify the relations that have been obtained from the FLR made into groups. Groups are created by matching the same relationships [5]. Example Group 1 (G1) have members  $A_1 \rightarrow A_1, A_3, A_7$  & Group 2 (G2) have members  $A_2 \rightarrow A_2, A_4, A_5$  etc.

#### 6. Defuzzification

The next step is to calculate the value of defuzzification, the way is by grouping carried out at the stage Fuzzy Logic Relations Group (FLRG). The step is to process all the members of the group that are joined and then the average is calculated. For example in Group 1 (G1) contain  $A_1 \rightarrow A_1, A_3, A_7$  then for  $A_1$  use  $u_1(m_1)$ ,  $A_3$  use  $u_3(m_3)$  and  $A_7$  use  $u_7(m_7)$  then calculate the average rata [6] with formula :

$$F_{(t)} = \frac{m_1+m_2+m_3}{3} \quad (5)$$

$m_1+m_2 + m_3$  = The middle value of the calculation from the initial stage

#### 7. Forecasting Calculations

The constancy of a good forecasting model can be calculated. Time series forecasting is good if it is close to the reality value. And also later the forecasting calculation can be seen in the results of the error the smaller the better [9].

#### b. Mean Absolute Percentage Error( MAPE)

MAPE is one of the most popular forecasting accuracy calculations used. MAPE has the following criteria:

**Table 2.** MAPE Criteria [5]

MAPE	Information
<10%	Very good forecasting ability
10% - 20%	Good forecasting ability
20% - 50%	Forecasting ability is sufficient
>50%	Poor forecasting ability

According to thesis [6], the following formula MAPE :

$$PE_i = \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \times 100 \quad (6)$$

$$MAPE = \sum_{t=1}^n \frac{|PE_i|}{n} \quad (7)$$

$\hat{y}_i$  = prediction value  
 $y_i$  = actual value  
 $n$  = number of data  
 $PE_i$  = percentage error

### c. Related Work

The research was conducted by Novi Ade Putra, Hendra Kurniawan and Nola Ritha, the purpose of this research is to see the accuracy of the algorithm used to assist the government in predicting population data in Tanjung Pinang City based on the smallest error value. One of them is that this research helps the government in the field of development and improves the standard of living for the welfare of the city. Because if an area has an increasing population, the higher the required investment. That way the researchers made predictions using the Fuzzy Time Series algorithm Chen's method and the measurement of the accuracy of the forecasting results using the Average Forecasting Error Rate (AFER). In this study, the variables used are month, year and the number of people who live in Tanjung Pinang City. The data that the researchers took was monthly from 2014 to 2016. According to the researchers, the Fuzzy Time Series algorithm Chen model could be used in solving the case of predicting the population of Tanjung Pinang City. The measurement error is obtained by an average value of 0.25% [8].

The research was conducted by Vivianti, Muhammad Kasim Aidid and Muhammad Nusrang. This study aims to forecast the number of visitors who travel to Fort Rotterdam, which is located in Makassar City, South Sulawesi. The tourism sector is one of the attractions that has benefits for the country in Indonesia. Currently the government has applied attention to developing the field of tourism, because the country of Indonesia has many unique features from its beautiful tourist attractions. In the field of tourism, this can produce an economy that has a good impact on Indonesia. And also get good foreign exchange earnings for the country and even the local community in general. Researchers predict that is by analyzing using time series data. The data was obtained from the Cultural Heritage Conservation Center (BPCB) which took 84 historical data, from January 2012 to December 2018. The researchers carried out forecasting by implementing two methods, namely the Chen model Fuzzy Time Series method and comparing it with the ARIMA method. Forecasting methods used by researchers are Mean Absolute Percentage Error (MAPE) and Root Mean of Square Error (RMSE). The forecast obtained for January 2019 is 16240.35. The forecast uses fuzzification in December 2018. The accuracy results are MAPE = 119.93 and RMSE = 4739.08 [5].

This study forecasts rainfall in Samarinda City for June 2016. Researchers predict using Chen's Fuzzy Time Series model. Researchers conducted three divisions of testing, including:

- a. January 2011 – May 2016 many as 65 data. Forecasting value of 268 mm by measuring the results of forecasting RMSE = 79,96 and MAE = 64,7.
- b. February 2013 – May 2016 obtained 287,5 mm and measurement of forecasting results RMSE = 84,91 and MAE = 64,4.
- c. January 2014 – May 2016 obtained 300 mm and measurement of forecasting results RMSE = 73,68 and MAE = 53,9.

From the test results which is the best number of samples in the period January 2014 – May 2016 as many as 29 data that produces the smallest error value compared to the other samples. [9].

### 3. Methodology

The type of data used in this study is secondary data. Source of data obtained at : <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>. Data files in the form of *time\_series\_covid\_19\_confirmed.csv*, the data taken is only the country of Indonesia. The data is in the form of chronological data regarding the development of the population of Covid-19 cases in Indonesia. Data is taken per day starting from 01 June 2020 – 29 May 2021, totaling 363 historical data as data that will be processed into this system research. The data needed is the variable of Time and Number of Population. Software development method that used in this research is Rational Unified Process(RUP)[10].

#### Software development method

#### Diagram Use Case

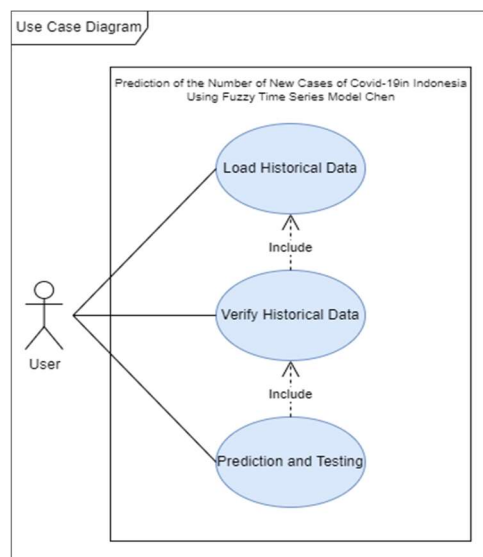


Figure 1. Use Case Diagram

Table 3. Use Case Definiton

No.	Use Case	Description
1.	Load Historical Data	User can enter a historical data file on the number of residents confirmed by Covid - 19 via the "Browse" button that is already available in the software.
2.	Verify Historical Data	User can display the verification value, which is the attribute value needed before forecasting from the historical data file that has been entered. Then the data will be able to go through forecasting calculation using Chen's Fuzzy Time Series model.
3.	Prediction and Testing	User can press the "Process" button to perform forecasting calculation and the percentage error of the forecasting that has been done. That way the user will know the value of the error rate. Calculation of the measurement of forecasting errors using MAPE

## 4. Result and Discussion

### i. Test data

The test is carried out by varying any positive number that has been determined by the researcher. Below is a table of the results of testing the researchers' calculations using the Chen FTSM method with input parameters  $D_1 = 135135$  &  $D_2 = 2000$ , as follows:

**Table 4.** Historical data

Date	Confirmed Quantity
01/06/2020	26940
02/06/2020	27549
03/06/2020	28233
04/06/2020	28818
05/06/2020	29521
06/06/2020	30514
07/06/2020	31186
..	..
28/05/2021	1803361
29/05/2021	1809926

**Table 5** Interval Determination

Class	Lower limit	Upper limit	Middle Value
A1	-108195	83817,1	-12188,95
A2	83817,1	275829,2	179823,15
A3	275829,2	467841,3	371835,25
A4	467841,3	659853,4	563847,35
A5	659853,4	851865,5	755859,45
A6	851865,5	1043877,6	947871,55
A7	1043877,6	1235889,7	1139883,65
A8	1235889,7	1427901,8	1331895,75
A9	1427901,8	1619913,9	1523907,85
A10	1619913,9	1811926	1715919,95

**Table 6.** Determine Fuzzy set, FLR and FLRG

Date	Confirmed Quantity	Fuzzification	LH	RH	FLR	FLRG
01/06/2020	26940	A1				
02/06/2020	27549	A1	A1	A1	A1->A1	G1
03/06/2020	28233	A1	A1	A1	A1->A1	G1
04/06/2020	28818	A1	A1	A1	A1->A1	G1
05/06/2020	29521	A1	A1	A1	A1->A1	G1
06/06/2020	30514	A1	A1	A1	A1->A1	G1

..						
..						
28/05/2021	1803361	A10	A10	A10	A10->A10	G10
29/05/2021	1809926	A10	A10	A10	A10->A10	G10

**Table 7** Defuzzification

Class	Current State	NextState	Defuzzification
A1	G1	A1, A2	83817,1
A2	G2	A2, A3	275829,2
A3	G3	A3, A4	467841,3
A4	G4	A4, A5	659853,4
A5	G5	A5, A6	851865,5
A6	G6	A6,A7	1043877,6
A7	G7	A7,A8	1235889,7
A8	G8	A8,A9	1427901,8
A9	G9	A9,A10	1619913,9
A10	G10	A10	1715919,95

**ii. The Chen Model Fuzzy Time Series Test Results and MAPE Results with different input parameters**

The test results were calculated using Chen's Fuzzy Time Series model. Below is table 8 which shows the results of prediction and forecasting accuracy in the experiment  $D_1 = 135135$  and  $D_2 = 2000$  and table 9 shows the MAPE value results for the influence test  $D_1$  and  $D_2$  with different input values.

**Table 8** Forecasting Test Value Results and Error Rate Value Results

Date	Confirmed Quantity	Deffuzification (Forecasting)	Percentage Error (PE)
01/06/2020	26940		
02/06/2020	27549	83817,1	204,2473411
03/06/2020	28233	83817,1	196,8763504
04/06/2020	28818	83817,1	190,8498161
05/06/2020	29521	83817,1	183,9236476
..			
21/12/2020	671778	851865,5	26,8075912
22/12/2020	678125	851865,5	25,62071889
23/12/2020	685639	851865,5	24,24402638
Total			12904,67467
Avarage			35,55006797%



Below, it is shown that there are 20 samples from 50 experimental influence tests with 363 data conducted by researchers to get a better MAPE value. There are columns for different input parameters  $D_1$  and  $D_2$  and MAPE values.

**Table 9.** MAPE Value Results for the influence test  $D_1$  and  $D_2$  with different values

No	$D_1$	$D_2$	MAPE
1.	0	20	62,03898584%
2.	1000	2000	61,76315575%
3.	50010,2	12000,4	48,10087735%
4.	50010,2	50010,2	49,67178113%
5.	50010,2	150010,2	53,5040072%
6.	150010,2	50010,2	37,50634577%
7.	250010	50010,2	60,0506414%
8.	250010	9010,2	56,98255909%
9.	170010	9010,2	40,22395592%
10.	140010	9010,2	35,76235541%
11.	140010	19010,2	36,22876567%
12.	130010	19010,2	35,83973888%
13.	230010	9010,2	62,94394373%
14.	140010	0	35,74964028%
15.	240010	0	58,73257626%
16.	240010	250	58,76287532%
17.	240010	16899	60,12773422%
18.	240010	168990	71,33291747%
19.	40010	168990	57,34829201%
20.	140010	168990	39,10869937%

The researcher analyzed that for the first positive number input or called parameter  $D_1$  has a good effect on the accuracy of the results compared to the parameter  $D_2$ . This parameter input has a good effect on achieving the desired accuracy value, because the interval at the lower limit greatly affects the forecasting (Defuzzification), the closer to the actual value, the better the accuracy. According to the researcher, the parameters  $D_2$  are only to help the parameters get a good accuracy value.

In conclusion, the results of the measurement of forecasting errors using MAPE in the results obtained the smallest error value of 35%. In the MAPE category, the value of 35% is in the sufficient category. According to researchers, the MAPE value can only reach sufficient because it depends on historical data. The historical data used by the researcher is data from patients who have confirmed Covid-19, which is increasing day by day, because these numbers affect the internal determination and grouping of the predicted values. This results in the prediction of historical data having a Percentage Error that is not yet close to the actual value. The closer to the actual value, the Percentage Error has a good forecasting ability.

## 5. Conclusion

The researcher implemented the Fuzzy Time Series logic in predicting Covid-19 cases in Indonesia by means of historical data in the form of the number of confirmed Covid-19 residents in Indonesia. The data is processed by performing several calculations, the first to determine the speaker universe as a division of determining intervals, determining attribute values such as Class Number, Class Length, Min Value, Max, etc. The second is to calculate the determination of fuzzy sets, FLR and FLRG. The third is doing defuzzification, namely forecasting.

Forecasting calculations using Chen's Fuzzy Time Series model, forecasting errors are calculated using the MAPE method. Researchers tested the effect with input parameters to get better MAPE results. Of 50 trials of parameter input with 363 data quantity, the result of the percentage inputted by the researcher is a positive number  $D_1 = 135135$  and  $D_2 = 2000$  very influential to get the most appropriate MAPE category. MAPE value result  $< 50\%$  show the average value *error* 35,55006797 (35%).

## References

- [1] Junaedi, D., & Salistia, F. (2020). Dampak Pandemi Covid-19 Terhadap Pasar Modal Di Indonesia: Al-Kharaj : Jurnal Ekonomi, Keuangan & Bisnis Syariah, 2(2), 109–138. <https://doi.org/10.47467/alkharaj.v2i2.112>
- [2] Agung, I. M. (2020). Memahami Pandemi Covid-19 Dalam Perspektif Psikologi Sosial. Psikobuletin:Buletin Ilmiah Psikologi, 1(2), 68–84. <http://ejournal.uin-suska.ac.id/index.php/Psikobuletin/article/view/9616/5058>
- [3] Ella, A. P., & Arifianto, D. (2020). Penerapan Fuzzy Time Series dalam Peramalan Harga Minyak Serai pada CV Agam Jaya Atsiri. 1–8.K. Elissa, “Title of paper if known,” unpublished.
- [4] Nugroho, K. (2016). Model Analisis Prediksi Menggunakan Metode Fuzzy Time Series. Infokam, 12(1), 46–50. <http://amikjtc.com/jurnal/index.php/jurnal/article/view/91>
- [5] Vivianti, V., Aidid, M. K., & Nusrang, M. (2020). Implementasi Metode Fuzzy Time Series untuk Peramalan Jumlah Pengunjung di Benteng Fort Rotterdam. 2(1), 1–12. <https://doi.org/10.35580/variansiunm12895>.
- [6] Febriana, E. T. (2018). Fuzzy Time Series Chen Orde Tinggi Untuk Meramalkan Jumlah Penumpang Dan Kendaraan Kapal (Studi Kasus: PT ASDP Indonseia Ferry (Persero) Cabang Merak). Universitas Islam Indonesia, Yogyakarta.
- [7] Saelan, A. (2009). Logika Fuzzy. Struktur Diskrit, 1(13508029), 1–5.
- [8] Ade Putra, N., Kurniawan, H., & Ritha, N. (2017). Prediksi Jumlah Penduduk Menggunakan Fuzzy Time Series Model Chen (Studi Kasus: Kota Tanjungpinang). Journal of Chemical Information and Modeling, 53(9), 1689–1699. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [9] Fauziah, N., Wahyuningsih, S., & Nasution, Y. N. (2016). Peramalan Menggunakan Fuzzy Time Series Chen (Studi Kasus : Curah Hujan Kota Samarinda). 4(2).
- [10] Perwitasari, R., Afwani, R., & Endang Anjarwani, S. (2020). Penerapan Metode Rational Unified Process (RUP) dalam Pengembangan Sistem Informasi Medical Check Up pada Citra Medical Centre (The Application of RationalUnifiedProcess(RUP)) in Development of a Medical. 2(1), 76–88.