Sentiment Analysis of the Indonesian Health Ministry Performance in Covid-19 Crisis using Support Vector Machine (SVM)

Fathiyarizq Mahendra Putra[a1], I Wayan Santiyasa[a2]

[a]Informatics Department, Faculty of Math and Science, Udayana University
Bali, Indonesia
[1]fathiyarizq.mahendra@gmail.com
[2]santiyasa67@gmail.com

Abstract

Corona Virus Disease or COVID 19 is a new virus disease that originated in 2019 [6], Indonesia has reported first COVID-19 In 2nd March 2020. Various attempts have been made by the government, such as taking strict measures by temporal lockdown or cordoning off the areas that were suspected of having risks of community spread. As a source of information, the internet has changed substantially,. for example, social media. social media is a communication tool that is very popular among internet users today, From social media, users can update status, send messages, even, become a platform for exchanging socio-economic opinions and political views both in their place of residence or their country. This paper deals with the sentiment analysis of Indonesian after the peformance of Indonesian Ministry Of Health. We used the social media platform Twitter for our analysis. Tweets were studied to gauge the opinion of Indonesian towards peformance of Indonesian Ministry Of Health. Tweets were extracted using the two prominent keywords used namely: "terawan "and "menkes" from June 15th to September 19th 2020. A total of 200 tweets were considered for the analysis. This study has successfully implemented the SVM algorithm for sentiment analysis on tweet data about peformance of Indonesian Ministry Of Health during COVID-19 Crisis. This is shown by the accuracy of using tweet data as much as 200 data, which is 172 data are training data and 28 are testing data. Besides the amount of data that affects accuracy, there are also other factors, namely the use of the kernel and the number of classes used. The results show that the Linear Kernel has the best accuracy, precision and recall rate compared to other kernels, respectively 75% for accuracy, 78.4% for precision and a recall value of 75%. for polynomial kernels, Gaussian and Sigmoid have the same accuracy, precision, and recall rates, namely, respectively. 60.71% for accuracy, 36.86% for precision and 60.71% recall value.

Keywords: *Sentiment, Analysis, SVM, COVID-19, Ministry Of Health, Indonesia*

1. Introduction

Corona Virus Disease or COVID 19 is a new virus disease that originated in 2019 [6], Indonesia has reported first COVID-19 In 2nd March 2020 [7]. Various attempts have been made by the government, such as taking strict measures by temporal lockdown or cordoning off the areas that were suspected of having risks of community spread [3]. Taking cues from the foreign counterparts, the government of Indonesia undertook an important decision to "flatten the curve." Such as Social Distancing in public places and province-wide lockdown.

Looking at the statistics of COVID19 infected, recovered, and death cases. Indonesians knew that drastic measures were needed in Indonesia to stop the numbers from rising exponentially, However, the weak decisions and poor performance taken by the Indonesian government, especially the performance of the minister of health, are one of the spotlight on social media because of the handling of COVID-19 which makes people still doubtful about the "flatten the curve." of Covid-19.

As a source of information, the internet has changed substantially from being a read-only source, to becoming a forum for interaction without borders and time. for example, social media. social media is a communication tool that is very popular among internet users today. Examples of social media are Facebook, Tumblr and Twitter, From social media, users can update status, send messages, even, become a platform for exchanging socio-economic opinions and political views both in their place of residence or their country, free message formats and accessibility from various devices are the main factors for internet users using social media.

Twitter is a Social Media site with more than 29.5 million users in Indonesia and 383 million tweets per day [1]. Twitter can be a great source of public opinion and sentiment data which can be used efficiently for marketing or social studies [5].

This paper deals with the sentiment analysis of Indonesian after the peformance of Indonesian Ministry Of Health. We used the social media platform Twitter for our analysis. Tweets were studied to gauge the opinion of Indonesian towards peformance of Indonesian Ministry Of Health. Tweets were extracted using the two prominent keywords used namely: "terawan "and "menkes" from June 15th to September 19th 2020. A total of 200 tweets were considered for the analysis. Analysis was done using the python and Support Vector Machine for sentiments classification of the tweets.

2.   Reseach Methods

Research Data

The data used in this study were taken from Twitter from two prominent keywords used namely: "terawan "and "menkes" from June 15th to September 19th 2020. A total of 200 tweets were considered for the analysis.

Data Processing

Data processing is carried out through the following process:

1. The preprocessing process carried out in this study through 4 stages, namely case folding, tokenizing, filtering or stopword removal and stemming. The preprocessing stage will produce a set of attributes or keywords in terms of terms (word for word).
2. In Case folding, this process are converting all the characters in a document into the same case, either all upper case or lower case.
3. The tokenizing stage. At this stage, certain characters such as punctuation are removed. While the space character is used as a delimiter to break a sentence into a collection of words
4. Filtering or Stopwords Removal are the words which does not add much meaning to a sentence. They can safely be ignored without sacrificing the meaning of the sentence. Then the words contained in the filtering process are converted into basic word forms using the stemming process
5. Weighting is done to get the value of the word / term that has been extracted. The most common method used to weight the term is TF-IDF weighting. The TF-IDF method is divided into integration between term frequency (TF) and inverse document frequency (IDF). The TF-IDF method can be formulated as follows:

$$TF * IDF = TF(w_i, d) * IDF(w_i) \tag{1}$$

Where :

w_i = word - i
d   = document
TF(w_i, d)   = he number of occurrences of the word w_i in the document d
IDF(w_i) = Invers Document Frequency of the word w_i

$$IDF(w_i) = log \frac{|D|}{DF(w_i)} \tag{2}$$

Where:

w_i = word - i
|D| = the total number of documents
IDF(w_i)   = the number of documents containing the word w_i

6. Classification using the Support Vector Machine (SVM) algorithm. The way SVM works is as follows:
   a. Data splits into training data and test data. Data sharing was done randomly. The first data to the 172 data are used as training data while the 173 to 200 data are used as test data. The training data is used for the training process in finding values a and b to be used in a decision function (classifier). While the test data is used for the testing process which will result in the classification of the data whether it is positive, neutral or negative. Data that states a positive tweet is labeled "positif" while data that states a neutral tweet is labeled "netral" amd data that states a negative tweet is labeled "negatif".
   b. Classification with SVM (Support Vector Machine) is looking for the best hyperplane that functions as a separator of two data classes [2]. SVM (Support Vector Machine) is able to work on high-dimensional datasets using kernel tricks. SVM (Support Vector Machine) is able to work on high-dimensional datasets using kernel tricks. SVM (Support Vector Machine) only uses a few selected data points that contribute (support vector) to form a model that will be used in the classification process
   Support Vector Machine Equation :

$$f(x) = w.x + b \qquad (3)$$

$$f(x) = \sum_{i=1}^{m} a_i y_i K(x, xi) + b \qquad (4)$$

   Where :

   w : Hyperplane parameter being searched for (the line perpendicular between the hyperplane line and the support vector point)

   x : Support Vector Machine input data points

   ai : the weight value of each data point

   $K(x,xi)$: kernel function

   b : hyperplane parameter to be sought (bias value)

   c. From the training process obtained parameter values α and b.
   d. After the training process is complete, the next testing process is to see the level of accuracy and error of the system. the error of the system. Generally, the measurement of classification performance is carried out with a configuration matrix. The confusion matrix is a table that records the results of classification work [6], The following table shows the confusion matrix for the two class classifiers listed in Table.1

Table 1. 3-Class Confusion Matrix

Actual	Predict		
	Negative	Neutral	Positive
Negative	TNg	NgN	FNg
Neutral	NNg	TN	NP
Positive	FP	PN	TP

In this study, entries in confusion matrix have the following meaning:
- FNg   = the number of incorrect predictions that the instances are negative
- TNg   = the number of correct predictions that an instance is negative.
- NgN   = the number of negative words that are predicted to be neutral
- NNg   = the number of neutral words that are predicted to be negative
- TN   = the number of correct predictions that an instance is neutral.
- NP   = the number of neutral words that predict positive
- TP   = the number of correct predictions that an instance is positive
- PN   = the number of positive words that predict is neutral
- FP   = number of incorrect predictions that an instance is positive

e. Based on the content of the confusion matrix, the precision, recall and accuracy values of the classification results can be found. To calculate accuracy, precision and recall the following formula is used:

f. Based on the content of the confusion matrix, the precision, recall and accuracy values of the classification results can be found. To calculate accuracy, precision and recall the following formula is used:

$$Accuracy = \frac{TNg}{TNg + NgN + FNg} \times 100 \tag{5}$$

$$Precision = \frac{TNg}{TNg + NNg + FP} \times 100 \tag{6}$$

$$F1 - Score = 2 \times \frac{(Recall * Precision)}{(Recall + Precison)} \tag{7}$$

3. Result and Discussion

The training data in this study amounted to 172 sentiment data that have been labeled by the author and used for the algorithm training process, Table 2 is an example of training data that has been labeled by the author.

Table 2. Example Of Training Data

No	Tweet	Label
1	Kalo Pak @jokowi, bener2 mau bereskan masalah kesehatan, copot dulu Terawan. Ganti dgn Menkes yg kapabel dan kompeten. Menkes skrg itu part of the problem, bukan part of the solution. Mau menyelesaikan masalah kok dgn pake masalah. Ya salah.. https://t.co/n7H4IXWR3V	Negative
2	Apresiasi pemerintah terhadap pahlawan tenaga kesehatan yang gugur dalam tugas menangani Covid19, kami bersama Menko PMK bapak Muhajir Effendy, Menkes bapak Terawan dan Kepala BNPB, bapak Dony Monardo memberikan penghargaan dan santunan kepada lima ahli waris di Lanud Hasanuddin https://t.co/szMtBOhoaU	Positive
3	Sy jd bertanya apakah Menkes Terawan itu seorang Dokter... ? Dokter Jihan: Pernyataan Menkes Terawan Penuh Keangkuhan #Covid-19 via @jpnncom https://t.co/vlHdOFRU9q	Neutral

The number of testing data used in this study is 28 and has been labeled the sentiment. This is done to evaluate system performance. Table 3 is an example of testing data.

Table 3. Example Of Testing Data

3.963 Artinya besok bisa saja tembus 4000 Situasi begini menkes republiknya masih saja bisa mengeluarkan statemen konyol Presiden @jokowi beri rakyat alasan, mengapa di situasi genting seperti ini Menteri Terawan tetap dipertahankan? Kayak ga punya "sense of crisis" gitu pak" https://t.co/V6oZYq2Fax

The preprocessing text includes the stages of case folding described in Table 4, Tokenizing in Table 5, Filtering in Table 6 and Stemming in Table 7.

Table 4. Case Folding

No	Tweet	Casefolding
1	Kalo Pak @jokowi, bener2 mau bereskan masalah kesehatan, copot dulu Terawan. Ganti dgn Menkes yg kapabel dan kompeten. Menkes skrg itu part of the problem, bukan part of the solution. Mau menyelesaikan masalah kok dgn pake masalah. Ya salah.. https://t.co/n7H4IXWR3V	kalo pak @jokowi, bener2 mau bereskan masalah kesehatan, copot dulu terawan. ganti dgn menkes yg kapabel dan kompeten. menkes skrg itu part of the problem, bukan part of the solution. mau menyelesaikan masalah kok dgn pake masalah. ya salah.. https://t.co/n7h4ixwr3v
2	Apresiasi pemerintah terhadap pahlawan tenaga kesehatan yang gugur dalam tugas menangani Covid19, kami bersama Menko PMK bapak Muhajir Effendy, Menkes bapak Terawan dan Kepala BNPB, bapak Dony Monardo memberikan penghargaan dan santunan kepada lima ahli waris di Lanud Hasanuddin https://t.co/szMtBOhoaU	apresiasi pemerintah terhadap pahlawan tenaga kesehatan yang gugur dalam tugas menangani covid19, kami bersama menko pmk bapak muhajir effendy, menkes bapak terawan dan kepala bnpb, bapak dony monardo memberikan penghargaan dan santunan kepada lima ahli waris di lanud hasanuddin https://t.co/szmtbohoau
3	Sy jd bertanya apakah Menkes Terawan itu seorang Dokter... ? Dokter Jihan: Pernyataan Menkes Terawan Penuh Keangkuhan #Covid-19 via @jpnncom https://t.co/vlHdOFRU9q	sy jd bertanya apakah menkes terawan itu seorang dokter... ? dokter jihan: pernyataan menkes terawan penuh keangkuhan #covid-19 via @jpnncom https://t.co/vlhdofru9q

Table 5. Tokenizing

No	Tweet	Tokenizing
1	Kalo Pak @jokowi, bener2 mau bereskan masalah kesehatan, copot dulu Terawan. Ganti dgn Menkes yg kapabel dan kompeten. Menkes skrg itu part of the problem, bukan part of the solution. Mau menyelesaikan masalah kok dgn pake masalah. Ya salah.. https://t.co/n7H4IXWR3V	['Kalo', 'Pak', 'jokowi', 'mau', 'bereskan', 'masalah', 'kesehatan', 'copot', 'dulu', 'Terawan', 'Ganti', 'dgn', 'Menkes', 'yg', 'kapabel', 'dan', 'kompeten', 'Menkes', 'skrg', 'itu', 'part', 'of', 'the', 'problem', 'bukan', 'part', 'of', 'the', 'solution', 'Mau', 'menyelesaikan', 'masalah', 'kok', 'dgn', 'pake', 'masalah', 'Ya', 'salah', 'https']
2	Apresiasi pemerintah terhadap pahlawan tenaga kesehatan yang gugur dalam tugas menangani Covid19, kami bersama Menko PMK bapak Muhajir Effendy, Menkes bapak Terawan dan Kepala BNPB, bapak Dony Monardo memberikan penghargaan dan santunan kepada lima ahli waris di Lanud Hasanuddin https://t.co/szMtBOhoaU	['Apresiasi', 'pemerintah', 'terhadap', 'pahlawan', 'tenaga', 'kesehatan', 'yang', 'gugur', 'dalam', 'tugas', 'menangani', 'kami', 'bersama', 'Menko', 'PMK', 'bapak', 'Muhajir', 'Effendy', 'Menkes', 'bapak', 'Terawan', 'dan', 'Kepala', 'BNPB', 'bapak', 'Dony', 'Monardo', 'memberikan', 'penghargaan', 'dan', 'santunan', 'kepada', 'lima', 'ahli', 'waris', 'di', 'Lanud', 'Hasanuddin', 'https']
3	Sy jd bertanya apakah Menkes Terawan itu seorang Dokter... ? Dokter Jihan: Pernyataan	['Sy', 'jd', 'bertanya', 'apakah', 'Menkes', 'Terawan', 'itu', 'seorang',

	Menkes Terawan Penuh Keangkuhan #Covid-19 via @jpnncom https://t.co/vlHdOFRU9q	'Dokter', 'Dokter', 'Jihan', 'Pernyataan', 'Menkes', 'Terawan', 'Penuh', 'Keangkuhan', 'via', 'jpnncom', 'https']

Table 6. Filtering

No	Tweet	Tokenizing
1	Kalo Pak @jokowi, bener2 mau bereskan masalah kesehatan, copot dulu Terawan. Ganti dgn Menkes yg kapabel dan kompeten. Menkes skrg itu part of the problem, bukan part of the solution. Mau menyelesaikan masalah kok dgn pake masalah. Ya salah.. https://t.co/n7H4IXWR3V	['Kalo', 'Pak', 'jokowi', 'mau', 'bereskan', 'masalah', 'kesehatan', 'copot', 'dulu', 'Terawan', 'Ganti', 'dgn', 'Menkes', 'yg', 'kapabel', '', 'kompeten', 'Menkes', 'skrg', '', 'part', 'of', 'the', 'problem', 'bukan', 'part', 'of', 'the', 'solution', 'Mau', 'menyelesaikan', 'masalah', 'kok', 'dgn', 'pake', 'masalah', 'Ya', 'salah', 'https']
2	Apresiasi pemerintah terhadap pahlawan tenaga kesehatan yang gugur dalam tugas menangani Covid19, kami bersama Menko PMK bapak Muhajir Effendy, Menkes bapak Terawan dan Kepala BNPB, bapak Dony Monardo memberikan penghargaan dan santunan kepada lima ahli waris di Lanud Hasanuddin https://t.co/szMtBOhoaU	['Apresiasi', 'pemerintah', '', 'pahlawan', 'tenaga', 'kesehatan', '', 'gugur', '', 'tugas', 'menangani', '', 'bersama', 'Menko', 'PMK', 'bapak', 'Muhajir', 'Effendy', 'Menkes', 'bapak', 'Terawan', '', 'Kepala', 'BNPB', 'bapak', 'Dony', 'Monardo', 'memberikan', 'penghargaan', '', 'santunan', '', 'lima', 'ahli', 'waris', '', 'Lanud', 'Hasanuddin', 'https']
3	Sy jd bertanya apakah Menkes Terawan itu seorang Dokter... ? Dokter Jihan: Pernyataan Menkes Terawan Penuh Keangkuhan #Covid-19 via @jpnncom https://t.co/vlHdOFRU9q	['Sy', 'jd', 'bertanya', '', 'Menkes', 'Terawan', '', 'seorang', 'Dokter', 'Dokter', 'Jihan', 'Pernyataan', 'Menkes', 'Terawan', 'Penuh', 'Keangkuhan', 'via', 'jpnncom', 'https']

Table 7. Stemming

No	Tweet	Stemming
1	Kalo Pak @jokowi, bener2 mau bereskan masalah kesehatan, copot dulu Terawan. Ganti dgn Menkes yg kapabel dan kompeten. Menkes skrg itu part of the problem, bukan part of the solution. Mau menyelesaikan masalah kok dgn pake masalah. Ya salah.. https://t.co/n7H4IXWR3V	['kalo', 'pak', 'jokowi', 'mau', 'beres', 'masalah', 'sehat', 'copot', 'dulu', 'terawan', 'ganti', 'dgn', 'menkes', 'yg', 'kapabel', '', 'kompeten', 'menkes', 'skrg', '', 'part', 'of', 'the', 'problem', 'bukan', 'part', 'of', 'the', 'solution', 'mau', 'selesai', 'masalah', 'kok', 'dgn', 'pake', 'masalah', 'ya', 'salah', 'https']
2	Apresiasi pemerintah terhadap pahlawan tenaga kesehatan yang gugur dalam tugas menangani Covid19, kami bersama Menko PMK bapak Muhajir Effendy, Menkes bapak Terawan dan Kepala BNPB, bapak Dony Monardo memberikan penghargaan dan santunan kepada lima ahli waris di Lanud Hasanuddin https://t.co/szMtBOhoaU	['apresiasi', 'perintah', '', 'pahlawan', 'tenaga', 'sehat', '', 'gugur', '', 'tugas', 'tangan', '', 'sama', 'menko', 'pmk', 'bapak', 'muhajir', 'effendy', 'menkes', 'bapak', 'terawan', '', 'kepala', 'bnpb', 'bapak', 'dony', 'monardo', 'beri', 'harga', '', 'santun', '', 'lima', 'ahli', 'waris', '', 'lanud', 'hasanuddin', 'https']
3	Sy jd bertanya apakah Menkes Terawan itu seorang Dokter... ? Dokter Jihan: Pernyataan Menkes Terawan Penuh Keangkuhan	['sy', 'jd', 'tanya', '', 'menkes', 'terawan', '', 'orang', 'dokter', 'dokter', 'jihan', 'nyata', 'menkes', 'terawan',

#Covid-19　　　　via https://t.co/vlHdOFRU9q	@jpnncom	'penuh', 'angkuh', 'via', 'jpnncom', 'https']

In this process, counting the number of terms or words that appear in a tweet (tf), counting the number of tweets containing that term (df), calculating the inverse document frequency (idf), and multiplying tf by idf as the weight of the terms in each tweet. Furthermore, sentiment analysis testing was carried out with 172 training data and 28 data testing, so that the total data amounted to 200 data. For the results of sentiment analysis using the SVM algorithm with the use of 200 data which is divided into 172 training data and 28 testing data evaluated by calculating the value of accuracy, precision and recall. Table 8 describes the confusion matrix of sentiment classification results.

Table 8. Confusion Matrix

Confusion Matrix			
	Negative	Neutral	Positive
Negative	16	0	1
Neutral	3	2	1
Positive	2	0	3

The SVM calculation results can also be influenced by the parameter selection. Parameters here can mean C values or kernel values. Comparison of kernel parameters. Linear Kernel, Polynomial Kernel and Radial Kernel or also called Gaussian Kernel and Sigmoid Kernel are listed in Table 9.

Table 9. Evaluation Of Kernel

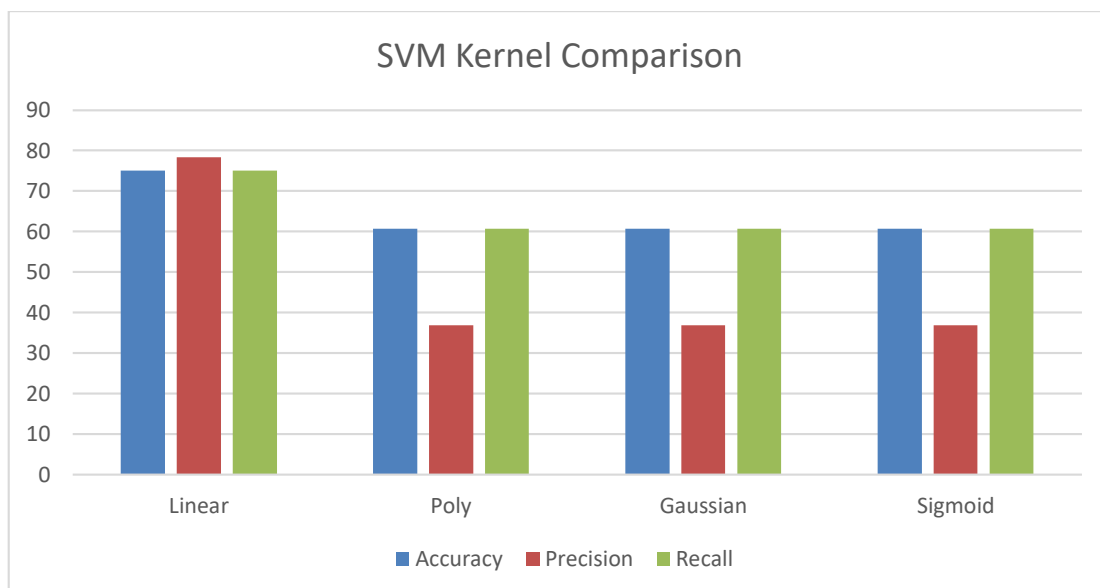|	Linear	Poly	Gaussian	Sigmoid
Accuracy	75.0	60.71	60.71	60.71
Precision	78.4	36.86	36.86	36.86
Recall	75.0	60.71	60.71	60.71

Figure 1. SVM Kernel Comparison

4.   Conclusion

This study has successfully implemented the SVM algorithm for sentiment analysis on tweet data about peformance of Indonesian Ministry Of Health during COVID-19 Crisis, using the two prominent keywords used namely: "terawan "and "menkes" from June 15th to September 19th 2020. This is shown by the accuracy of using tweet data as much as 200 data, which is 172 data are training data and 28 are testing data. Besides the amount of data that affects accuracy, there are also other factors, namely the use of the kernel and the number of classes used. The results show that the Linear Kernel has the best accuracy, precision and recall rate compared to other kernels, respectively 75% for accuracy, 78.4% for precision and a recall value of 75%. for polynomial kernels, Gaussian and Sigmoid have the same accuracy, precision, and recall rates, namely, respectively. 60.71% for accuracy, 36.86% for precision and 60.71% recall value.

References

[1]   Carley, K., Malik, M., Kowalchuck, M., Pfeffer, J., & Landwehr, P. (2015). *Twitter Usage in Indonesia.*

[2]   Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

[3]   Indonesia, P. (2018). Undang-Undang Nomor 6 Tahun 2018 tentang Kekarantinaan Kesehatan. *Jakarta: Kementerian Sekretariat Negara Republik Indonesia*.

[4]   Nomleni, P. (2015). SENTIMENT ANALYSIS MENGGUNAKAN SUPPORT VECTOR MACHINE(SVM) . SURABAYA: PROGRAM PASCA SARJANA BIDANG KEAHLIAN TELEMATIKA (CIO) JURUSAN TEKNIK ELEKTRO FAKULTAS TEKNOLOGI INDUSTRI INSTITUT TEKNOLOGI SEPULUH NOPEMBER .

[5]   Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning. *arXiv preprint*.

[6]   Wang, H., Wang, Z., Dong, Y., Chang, R., Xu, C., Yu, X., . . . others. (2020). Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China. *Cell discovery*, 1--8.

[7]   Yulisman, L. (2020, MAR 2). *Mother and daughter test positive for coronavirus in Indonesia, first confirmed cases in the country.* Retrieved from The Straits Times: https://www.straitstimes.com/asia/se-asia/indonesia-confirms-two-coronavirus-cases-president