Smart Environment Monitoring and Analytic in Real-time System for Vannamei Shrimp

Sritrusta Sukaridhoto[1,3], Dwi Kurnia Basuki[1], Jordan Maulana[1], Riyadh Arridha[2], and Tiyo Avianto[3]
[1]*Politeknik Elektronika Negeri Surabaya, Indonesia*
[2]*Politeknik Negeri Fakfak, Indonesia*
[3]*Eyro Digital Teknologi, Surabaya, Indoneisa*
dhoto@pens.ac.id

***Abstract*—dWater quality is the most important thing in shrimp farming to keep the shrimp healthy and well grown. Embedded system and sensors can be applied to monitor water quality for shrimp. This paper describes a Smart Environment Monitoring and Analytic in Real-time (SEMAR) platform that able to monitor pH, Dissolved Oxygen (DO), and temperature in Water Pond. Then the measured values are stored in Big Data system and analyzed with multi class classification of shrimp water quality. The system is able to send early warning for bad water quality and show real-time monitoring data by internet and smart phone.**

***Index Terms*— vanamei shrimp, monitoring, analytic, internet of things**

I. Introduction

Most of Indonesian shrimp farmer are still using traditional method on shrimp farming. In general, shrimp farmers only feed the shrimp, they don't care about water quality nor use the equipment to make shrimp live better. Shrimp farmers, by this way, control water quality manually by looking at shrimp behavior. Sometimes they get much, but sometimes they get few. However, shrimp is one of the Indonesian biggest export commodity. In 2015 only, Indonesia exported totally 145,077.9 tons of shrimp and gained 1,311,010,900 USD from it [1].

There are 4 water parameters that have big impact to vannamei shrimp [2], those are: salinity, dissolved oxygen (DO), water temperature, and potential of hydrogen (pH). Those parameters are often change and negatively impact to shrimp if they overlap shrimp's water quality tolerance. Vannamei shrimp can live in freshwater, so salinity is not used to classify water quality.

SEMAR (Smart Environment Monitoring and Analytic in Real-time System) is an ongoing research about big data technology for real-time analysis. SEMAR consist several joined research about environment monitoring to develop one big integrated system.

Decreasing dissolved oxygen, unstable pH and temperature value are lethal for shrimp [3]. At the same time salinity and turbidity support shrimp's life. In implementation, water quality values will be fetch by DO, pH, and temperature sensors from Atlas Scientific. The data then processed by raspberry and uploaded to physical server through MQTT. The data received by server will be classified as pond's water quality that can be seen from web and android application. If the data is classified as bad or lethal, the system immediately push a notification to the farmer's phone.

II. Related Work

Atlas scientific water sensors combined with radio-controlled submarine was used to collect information in underwater. The submarine contains IMU, GPS, micro controller and mini pc. The system was deployed well in real environment [4].

Low-cost, efficient, low-power water monitoring system was built from low power embedded system. As a hand-held device, it contains Raspberry Pi. Touch screen LCD, atlas scientific sensor kit and 3G USB modem [5].

New form of implementation of water monitoring system was analyzing the parameters to predict what causing it and understanding the impact or damage to living organism along the device was deployed [6].

In 2010, researchers from NECTEC-ACE have conducted study of Dissolved Oxygen level monitoring in shrimp aquaculture using embedded system. The system consist of Solar cell, MCU, IO module, Ethernet gateway and GPRS module, also aerator in form of paddlewheel [7].

Richter (2015), have conducted comparative studies in big data machine learning toolkits including Mahout MapReduce, Mahout Samsara, Spark MLlib, H2O, and SAMOA. The study concluded that in general, Spark MLlib and H2O have better performance than other toolkits in terms of extensibility, scalability, usability, fault tolerance and speed. SAMOA implements 3 algorithms, Mahout Samsara does 7, H2O does 10, Mahout MapReduce does 13 and Spark MLlib does 17 algorithms. Spark MLlib has additional advantage. It has an ability to cover batch and stream processing. Mahout and H2O only cover batch processing, and SAMOA only covers stream processing [8].

Katherin Indrawati (ITS, 2008), conducted study to create control module to maintain water quality. In this research, water pond is presented by miniplant. Monitored water quality are pH, salinity and temperature. Aerator is used to maintain water temperature and pH, meanwhile water pump is used to maintain salinity. The controller uses ATMega to receive data and activate the aerator and water pump [2].

Goib Wiranto (LIPI, 2010), conducted a study to develop water quality monitoring system for shrimp water pond. The system monitors DO and pH, uses SMS gateway and data logger. Sensor Lutron is used to collect pH and DO value and PLC is used as the controller. In this system, if any value exceeds shrimp's health, controller will send sms to user by sms gateway with GSM modem [9].

Rizqi et al., have also developed Atlas Scientific sensors to monitor river water quality as surveillance of liquid waste pollution. Water quality that is monitored are DO and pH [10].

III. SYSTEM DESIGN

Fig. 1 shows the IoT reference model. Our system design was based on the IoT reference model which consists of seven sections: 1) physical devices and controllers, 2) connectivity, 3) edge computing, 4) data accumulation, 5) data abstraction 6) application and 7) collaboration and processes. Next, Fig. 3 shows the overall system design. Physically, the infrastructure of this system consists of one sensors, Raspberry Pi 3 and physical server.

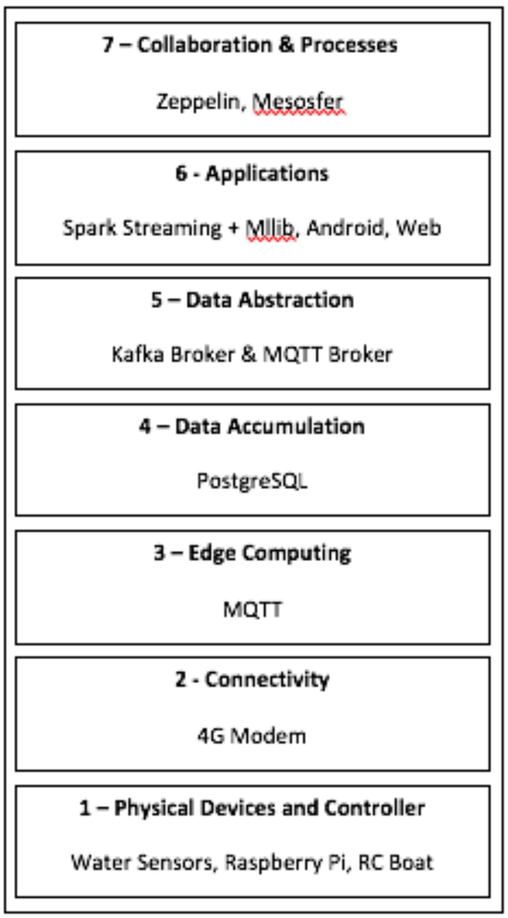
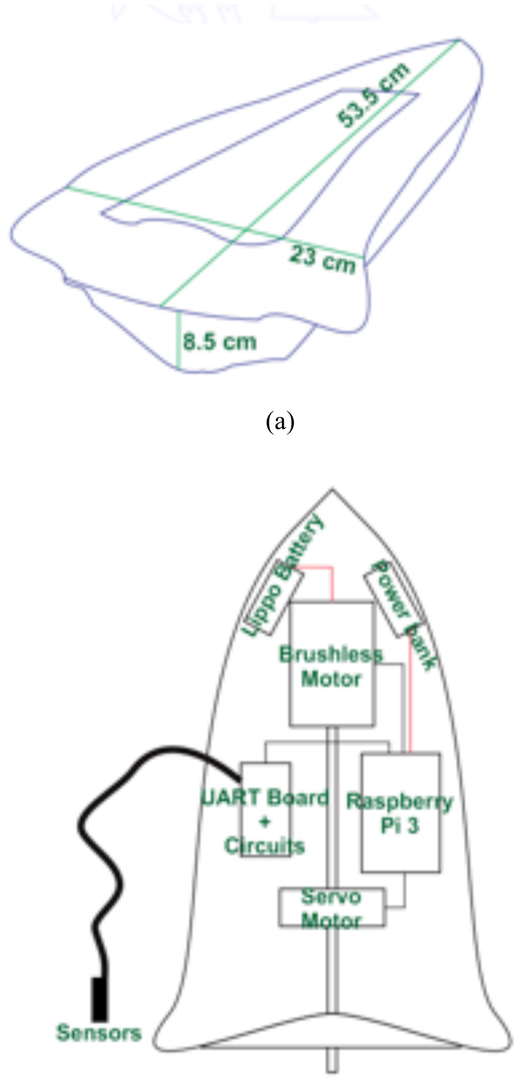


Figure 1. IOT Reference Model

A. Physical Devices and Controllers

Physical Devices and Controllers in this paper contain sensors, mini pc as the main controller and RC Boat. Water quality sensors come from 'Atlas Scientific' kit water sensor. The sensors include pH (Potential of Hydrogen), DO (Dissolved Oxygen), and temperature. In this research, Raspberry Pi 3 type B is used as Mini PC. Raspberry collects the values of sensors and combine them into 1 string with this format: pH, DO, temp. RC boat contains brushless motor and servo motor. Brushless motor is used as throttle and servo motor is used as steering. As the boat is working, we can collect data from different spots. Fig. 2 shows all the hardware of physical devices and controllers. Table 1 shows hardware details and its cost.



(b)

Figure 2. System Hardware

B. Connectivity

4G modem is used to connect Raspberry will connect to this Device through wireless connection.

C. Edge Computing

The process of data acceptance in the server uses MQTT broker by Mosquitto. A line of water quality data is received through 'tambakmonitoring' topic and stored to database.

D. Data Accumulation

Data accumulation accumulation is a process of storing water data from sensors to the PostgreSQL database table. Spark Streaming consumes/subscribes water data from Kafka Broker on topic 'watermonitoring'. There is a table in PostgreSQL database. The table named 'logair'. This table is used to store water quality data when the water quality changes. The table consists of 5 columns. They are pH, DO, temp, quality, time.

E. Data Abstraction

Platform that is used to manage the data flow on big data server is Kafka Broker. After water quality data is received by MQTT broker, it is forwarded to Kafka broker to arrange data flow. MQTTKafkaBridgeTambak is used to distribute data from MQTT broker to kafka broker. Then. Kafka Broker publishes the data to subscriber applications. Kafka Producer stands for data sender, Kafka Consumer stands for any application that subscribes it and Kafka Broker stands as an intermediary. Topic of MQTT to receive incoming data from mini pc is 'tambakmonitoring', while topic of Kafka Broker to publish data is 'tambakanalytic'.

F. Applications

The application layer of this SEMAR extension consists of: 1) Learning Process, 2) Real-time classification, 3) Real-time visualization in web and android, 4) Early warning in form of android notification.

G. Collaboration and Processes

Zeppelin is used to provide interface of the interaction between analyst and system. Classification model can be performed through web by integrating Zeppelin with Apache Spark. Mesosfer is also used as push notification platform. Mesosfer is an IoT platform and Mobile Backend as a Service which enable user to integrate and cut down development process.

IV. PROPOSED SYSTEM

A. Water Quallity Parameters

Each parameter monitored in this research is crucial for vannamei shrimp.

- Potential of Hydrogen (pH)

Normal pH value for shrimp is between 6 and 9. pH above 10 is lethal for shrimp and pH below 5 causes shrimp's growth slowing down. Specially, for vannamei shrimp, good pH value is between 7.5 and 8.5.

- Dissolved Oxygen (DO)

Oxygen is needed for shrimp to breath. A good oxygen for shrimp values above 3 ppm (part per million). It is recommended to set the values to 4 and 8 ppm. Low dissolved oxygen in the water often occurs on windless summer season. Besides, at night, temperature becomes low followed by high activity of Phytoplankton. This low oxygen condition can be shown by surfacing shrimp.

- Temperature

Temperature is one of the defining factors for shrimp's life. Good temperature for shrimp values between 26^0C - 30^0C. If temperature drops below 25^oC, digestibility of the shrimp will be reduced. If temperature rise above 30^oC, shrimp will be stressed in cause of high oxygen needing. Meanwhile, if temperature drops below 14^oC, shrimp will be dead

B. Rule Building

Table 1 show parameters value that influence shrimp's life. From the facts of water quality parameters given, we built a rulings to classify water quality for vannamei shrimp. It is shown in table 2.

Table 1. Parameter Value that influence in the shrimp's life

Params	Optimal Value	Bad Value	Lethal Value
Temperature (Celcius)	26-30	<25 & >30	<14
Potential of Hydrogen (pH)	7.5-8.5	<5	>10
Dissolved Oxygen (ppm)	>3	<3	<1

Table 2. Classification Rule

Water Quality is classified as 'Optimal' if all parameters value as Optimal, classified as 'Bad' if any parameters have bad value and classified as 'Lethal' if any parameters have lethal value. Real data from shrimp water pond is used to be the dataset. The data was collected within 2 days. The dataset will be used for training and forming classification model.

Table 3. Dataset Sample

pH	DO	Temp	Class
8.848	18.59	30.11376	1
8.848	19.29	30.11376	1
8.849	17.7	30.85104	1
8.849	18.41	30.11376	1

Table 3 shows a sample of approximately 5000 data that was collected on Keputih pond. Class 2 is representation of lethal, class 1 for bad, and class 0 for optimal water quality for vanamei shrimp. In 2 days of collecting data, the water quality was almost completely in bad condition.

C. Training Process

Training process of the dataset is conducted using Spark MLlib. Spark SQL makes data analysis able to be performed on large scale. Classification has several algorithm that can be supported by Spark MLlib. We use two classification algorithms, where the results would be compared to choose the best one. Those classification algorithms are Support Vector Machine and Decision Tree.

The hold-out method is used to evaluate the accuracy of classifier. This method divides the dataset into training set and test set. Training set is a subset that is used to construct classification model, and test set is a subset that is used to measure the performance of built classification model. 70% of the dataset is used for training, and the remaining 30% is used for test set. Learning procedure is shown in algorithm 1.

Algorithm 1 Learning Procedure.

```
Method ← Linear SVM, Decision Tree
```

```
Begin
  sensorInfo ← pH, DO, temp
  loop:
    Retrieve Dataset
    Split Dataset (70,30)
    Machine Learning Training (method)
    Machine Learning Testing (method)
    Calculate MSE, Mislabel, Accuracy
    Save model
  End
```

D. Real-time Classification

Big data analytic technology is used for real-time classification. High speed processing is needed in order to complete this task. It must also be stand-alone application in order to cut the delay between data received until the visualisation. Spark MLlib is used both for real-time classification and learning process. Spark reads water data from Kafka Broker by topic 'tambakmonitoring' by streaming, then conduct classification which produces the prediction of water quality for shrimp. Variable 'label' is used to save classification result. This label has numerical type and it is retranslated into categorical data (Optimal, Bad, Lethal) stored in new variable called 'criteria'. Kafka Broker then loads the result of classification by using Kafka Producer on port 9092 with the topic 'tambakanalytic'. This topic is used for real-time visualisation. Data is sent in form of JSON format with arrangement data of ID, Latitude, Longitude, Date, Time, pH, EC, TDS, Sal, DO, Temperature, Depth, Label, Criteria. The algorithm of the real-time classification can be seen on algorithm 2.

Algorithm 2 Real-time Classification.

```
function parsing(data)
    parse ← param1, param2, …, param n
    return parse
end function
function realtimeClassification(parse)
    classify ← Model.predict
    return classify
end function
function sendToKafka(data,classify)
    Send data, classify to Kafka Broker
end function
Begin
    Spark initialization
    Load Model
    data ← pH,DO,Temp from Kafka Broker
    parsing (data)
    realtimeClassification (parse)
    SendToKafka(data,classify)
    End
```

E. Visualization

The data published by Kafka Broker is read by Node JS using Kafka Consumer with the topic 'tambakanalytic'. Node JS makes the flow of real-time data able to be managed to the front end. Node JS uses web socket to send the data from back end to front end. Front end layer consists of Apache, PHPCI, Javascript, CSS and HTML. Google Map API is used to visualize the location of sensor node and Highchart is used to visualize water data charts. Water data can also be shown in android application. Fig. 4 shows the visualization of the web interface. The web contains blocks to select parameter that is wanted to be shown, followed by chart and water quality classification. Fig. 5 shows the visualization of monitoring, one of the features available in the android app. The features of android app are 1) show and update user profile. 2) show whoever installed the app. 3) show system log. 4) show present water quality data in real-time. 5) control the ship.

V. EXPERIMENT RESULT

The experiments were conducted by testing the performance of linear SVM and Decision Tree algorithm using their default parameters. We used dataset from randomly generated number corresponding to the rule given.

A. Confusion Matrix Results

Table 1 and 2 show the confusion matrix of training model from Linear SVM and Decision Tree.

Table 4. Linear SVM Dataset Training Linear Support Vector Machine

True Label	0	0	5	0
	1	0	1536	2
	2	0	0	143
		0	1	2
			Predicted Label	

The SVM algorithm results some wrong prediction significantly. Every label has high number of wrong prediction label.

Table 5. Decision Tree Dataset Training

True Label	0	4	1	0
	1	20	1536	2
	2	0	0	143
		0	1	2
			Predicted Label	

Different from SVM, Decision tree has higher accuracy of predicted label. So, with this comparement, we choosed Decision Tree algorithm to be used for machine learning as it has high accuracy of prediction.

B. Classification Results

The validation of classification model can also be shown with ROC (Receiver Operating Curve). ROC compares graph between TPR (True Positive Rate) on vertical axis with FPR (False Positive Rate) on horizontal axis of the ROC. The area under ROC curve is known as AUC (Area Under ROC Curve). AUC value ranges from 0 to 1. Value means better when it gets closer to 1. From conducted experiments, SVM shows 72% of average accuracy from all classes. Decision Tree has average accuracy of 96%. The graphical results of experiment is shown in Figure 4 for SVM and Figure 5 for Decision Tree.

ROC in Fig. 6 shows that Support Vector Machine algorithm has performance of $0.78 \leq AUC \leq 0.97$ and Fig. 7 shows ROC of Decision Tree algorithm.

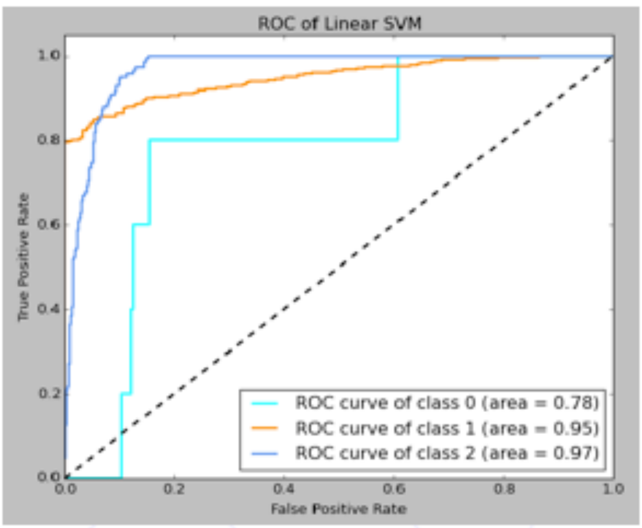


Figure 3. ROC of Dataset Using Linear Support Vector Machine

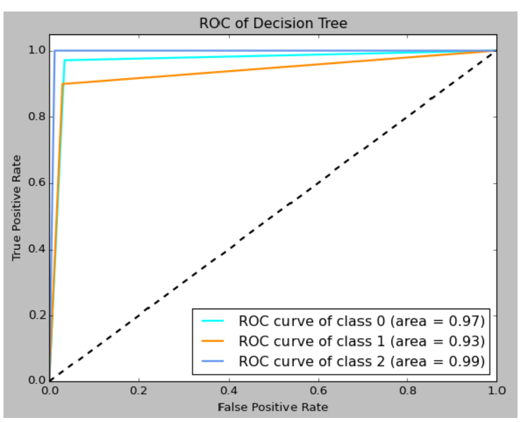


Figure 4. ROC of Dataset Using Decision Tree

Different from SVM, Decision Tree has much higher accuracy on all 3 classes that values $0.97 \leq AUC\ 0.99$. So Decision Tree is used instead of SVM.

C. Real System Implementation

The experiment was done by comparing machine learning data with manually classification data. First, Raspberry collects water quality data, then save it in CSV file format. At the same time, it uploads the data to the server. As machine learning system is running, it collects uploaded data and classifies them. The result of its classification is saved in CSV file format. Comma-separated values (CSV) file stores tabular data (numbers and text) in plain text. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. The use of the comma as a field separator is the source of the name for this file format.

Table 6. Comparation of Manual and Machine Learning Classification In Real Water Pond

Location	Method	Classification			Error	Push
		Optimal	Bad	Lethal		
1	Manual	-	1350	2	0%	4
	Learning	-	1350	2		
2	Manual	156	37	-	0.51%	30
	Learning	157	36	-		
3	Manual	237	42	-	0.71%	35
	Learning	239	40	-		
4	Manual	319	40	-	0.50%	33
	Learning	321	38	-		

Table 6 shows that at the time the experiment was conducted, the shrimp water pond is at bad condition. Then the system immediately do its early warning feature, push notification to android app. The activity is shown in Fig. 8 and Fig. 9 shows the implemented hardware in water pond.

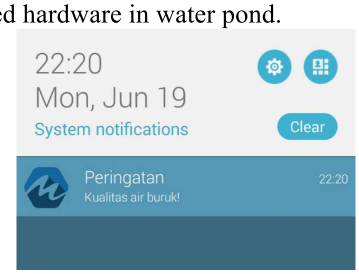


Figure 5. Early Warning Push notification to android app.

VI. CONCLUSION

In this paper, the new designs of shrimp monitoring was tested great. Firstly, the water quality data is fetched through water quality sensors. Then, it is uploaded to the server by MQTT protocol. The server runs classification based on training model. Decision Tree method gives precise classification that its error is less than 1%. The result of data monitored and its classification can be seen on web and android smartphone.

REFERENCES

[1] Ekspor Udang Menurut Negara Tujuan Utama 2000-2015, https://www.bps.go.id/index.php/linkTabelStatis/1015, May 2017.

[2] Katherin Indriawati, “Pembuatan modul kontrol kualitas air tambak udang sebagai sarana pembelajaran perbaikan teknik budidaya udang”, 2008.

[3] Khairul Amri, and Iskandar Kanna. Budidaya Udang Vaname. N.p.: Gramedia Pustaka Utama, 2013. Print.

[4] Sukaridhoto, Sritrusta, Dadet Pramadihanto, Muhammad Alif, Andrie Yuwono, and Nobuo Funabiki. "A design of radio-controlled submarine modification for river water quality monitoring." In Intelligent Technology and Its Applications (ISITIA), 2015 International Seminar on, pp. 75-80. IEEE, 2015.

[5] Sukaridhoto, Sritrusta, Rahardhita Widyatra Sudibyo, Widi Sarinastiti, Rizky Dharmawan, Atit Sasono, Ahmad Andika Saputra, and Shiori Sasaki. "Design and development of a portable low-cost COTS-based water quality monitoring system." In Intelligent Technology and Its Applications (ISITIA), 2016 International Seminar on, pp. 635-640. IEEE, 2016.

[6] Rindang, Tegar Esa. "Analisis dan pemetaan data parameter kualitas air dengan menggunakan open rov.", 2016.

[7] “Dissolved oxygen level monitoring in shrimp aquaculture using embedded system”. N.p.: Nectec, 23 Sept. 2010. PDF.

[8] Richter, Aaron N., Taghi M. Khoshgoftaar, Sara Landset, and Tawfiq Hasanin. "A multi-dimensional comparison of toolkits for machine learning with big data." In Information Reuse and Integration (IRI), 2015 IEEE International Conference on, pp. 1-8. IEEE, 2015.

[9] Goib Wiranto, and I. Dewa Putu. “Pembuatan sistem monitoring kualitas air secara real time dan aplikasinya dalam pengelolaan tambak udang” (2010).

[10] M. Farkan, D. D. Setiyanto, R. S. Widjaja, Kholil, and Widiatmaka. "IOPscience." Assessment Area Development of Sustainable Shrimp Culture Ponds (case Ctudy the Gulf Coast Banten) (2017): n. pag. Web. 9 Aug. 2017.

[11] Goib Wiranto, Yudi Yuliyus Maulana, Dewa Putu Hermida, Iqbal Syamsu, and Dadin Mahmudin. Ntegrated Online Water Quality Monitoring An Application for Shrimp Aquaculture Data Collection and Automation (2015): n. pag. IEEE Explore Digital Library. 12 Nov. 2015. Web. 14 Aug. 2017.

[12] Baharuddin Aziz, Elvayandri Muchtar, and Farkhad Ihsan Hariadi. Human-Machine Interface for Water Quality Monitoring System of White-Legged Shrimp Pond (2017): n. pag. IEEE Explore Digital Library. 27 Mar. 2017. Web. 14 Aug. 2017.