Risk Assessment Based on Hierarchical Fuzzy Inference and Prediction using Kalman Filter for Underground Facilities in Smart Cities

[1]Israr Ullah, Muhammad Fayaz, DoHyeun Kim

Abstract: Background/Objectives: In developed world, the trend of underground facilities construction is on the rise as this provide a way to maximum utilization of scarcest resource in smart cities i.e. space. However, these facilities can cause serious damage to property and human lives if not monitored properly. This paper presents a risk prediction and assessment mechanism using Kalman filter and hierarchical fuzzy inference for risk analysis of underground facilities. Methods/Statistical analysis: Sensors are installed at selected locations to collect input data by observing underground facilities. After necessary data processing, sensor data is used to compute risk index of individual factor that ultimately contribute toward total risk index of underground facility. For computing total risk index from given input factors, we use hierarchical fuzzy logic-based system by combining related components in a tree like structure. Experiments are conducted with four selected factors to estimate total risk index of underground facilities. Every risk factor is estimated from underground sensors data which may have noise and error. Noise and error from input data is removed using Kalman Filter algorithm. Findings: Results shows that without applying Kalman filter prediction, results for estimation of final risk index are not satisfactory. Kalman filter prediction helps in removal of noise in sensor reading and provides an effective way for the monitoring of underground facilities by forecasting critical issues in advance to avoid potential damage and can also assist in improving the maintenance work. Improvements/Applications: In this paper, we have used simulation data. In future, we will use real data to verify the proposed model for underground risk prediction and assessment.

Keywords: Risk prediction, Kalman filter, Risk assessment, Hierarchical fuzzy inference system.

I. INTRODUCTION

Every decision-making process demands a clear understanding risks trends in future or otherwise losses can be increased [1]. Therefore, many people tend to delay the decision as much as possible such that satiation gets clear to make any decision. However, delaying the decision is never a wise approach in current competitive market conditions. Therefore, many prediction algorithms are proposed to learn from history and make intelligent about future conditions to support early decision-making process [2]. We can make intelligent guess about future using current data combined with history data and thus better decision can be made in advance to maximize profit or avoid loss [3].

These days, almost every scientific discipline make use of prediction algorithm in one way or the other. Study of machine learning algorithms have tremendously evolved in recent past due to significant advancement in computational capabilities of computers. To predict, predictive modeling is used. Predictive modeling uses statistics to predict outcomes [4]. Most often the event one wants to predict is in the future, but predictive modelling can be used for any type of unknown event, regardless of when it occurred. For prediction purposes nearly, any regression model can be used. Numerous algorithms use for prediction purposes, such as K-nearest neighbor algorithm(KNN) [5], Support vector machines [6], Random forests [7], neural network [8] etc. We have selected Kalman Filter algorithm because of its simplified and lightweight design that suits well to our requirement [9].

In developed world, the trend of underground facilities construction is on the rise and many countries are adopting this solution to overcome the space shortage in smart cities [10]. Among such underground facilities includes, the parking areas, railway tracks, water supply and electric lines, sewer management etc. [11-12]. Although these facilities provide benefit in terms of space utilization, but the associated risk factors and threats to human lives and properties often become hindrance in worldwide adaptation of this solution. To overcome the risk factors and avoid potential threats, we need to develop a real time monitoring system that can continuously conduct risk assessment to timely report any abnormal conditions that can lead to sewer damage [13]. This work presents an IoT based system for automated risk analysis and prediction.

II. MATERIALS AND METHODS

Conceptual design of proposed system is given in Figure 1. First, we collect data about the various risk factors through sensors. Usually, data collected from sensor is not clean and often noise from various sources generates spikes in data. Many different methods are developed to remove noise and estimate actual data point in data. Among those methods, Kalman's filter is very popular due its light weight nature and intelligent mechanism to predicted system actual state under noisy conditions. Its light weight as it does not require lots of history data to predict or estimate system actual state, rather only system previous state is required. Kalman filter algorithm is used in this study to remove noise from sensing data collected for

Revised Manuscript Received on January 03, 2019.

Israr Ullah, Computer Engineering Department, Jeju National University, Republic of Korea,

Muhammad Fayaz, Computer Engineering Department, Jeju National University, Republic of Korea,

DoHyeun Kim, Corresponding author, Professor, Computer Engineering Department, Jeju National University, Republic of Korea

various risk contributing factors under study. After, removing the noise from risk contributing factors, we have used hierarchical fuzzy inference system to estimate final risk index for given location using various associated risk contributing factors data. Conventional fuzzy inference system (FIS) have the problem of rules explosion with increase in input parameters which is resolved in hierarchical fuzzy inference system where inputs are grouped and processed in multiple layered architecture. The whole system consists of FIS components connected in a tree-like structure. As per relationship among the input parameters (which is often unknown), the tree structure can be changed.

Figure 1. Proposed Risk Assessment Mechanism Based on Hierarchical Fuzzy Inference and Prediction

2.1. Risk Prediction using Kalman's Filter Algorithm

As stated earlier, sensing data often catch noise from various unknown sources which effect the accuracy of the system using that noisy data. Kalman's filter algorithm is used in this study to remove noise from sensing data. Kalman filter require only the information about system previous state to predict its next state using estimated noise in sensing data and the expected error in the system model. Kalman filter algorithm can converge quickly irrespective of the initial/starting state guess of the system. At the core of this algorithm is a parameter known as Kalman's gain (K) which perform all the magic. This K define and decides whether more preference should be given system model or sensor readings to predict system future state. Figure 2 presents working of Kalman filter algorithm. In our case, sensor reading is the estimated risk index values for individual risk contributing factors and we have assumed error in sensor reading with certain probability. Kalman's filter algorithm is used to remove error and noise from sensing data and thus useful in predicting actual risk index with high accuracy.

2.2 Mathematical formulation for predicted risk index using Kalman's Filter

Let's assume location (x, y) has the risk index Risk_tat time 't'. Using estimated risk index Risk_t, we can estimate risk index at t+1 i.e. Risk_{t+1}. Steps included in Kalman filter algorithm are as follows.

First, predicted risk is computed from estimated risk at previous time t-1 i.e.

$$\mathrm{Risk}_{predicted} = A.\mathrm{Risk}_{t-1} + B.u_t \quad (1)$$

Here, A express the state transition matrix, matrix B express the process control matrix, Risk_{t-1} is risk at time t-1 and control vector is denoted with u_t.

Next, predicted covariance factor is updated as below

$$P_{predicted} = A.P_{t-1}.A^T + Q \quad (2)$$

where A is the same state transition matrix and A^T is its transpose, P_{t-1} is the process covariance calculated previously and process error is estimated using Q.

After above necessary formulation, next we update the Kalman's gain (expressed as K) using following equation.

$$K = \frac{P_{predicted}.H^T}{H.P_{predicted}.H^T + R} \quad (3)$$

where H is observation matrix and H^T are its transpose. Measurement error is expressed as R.

Let's assume, our risk calculation from current sensor's reading using hierarchical fuzzy logic system at time "t" is z_t. Then, predicted risk index for current time interval using Kalman's filter can be computed as below.

$$\mathrm{Risk}_t = \mathrm{Risk}_{predicted} + K(z_t - H.\mathrm{Risk}_{predicted}) \quad (4)$$

Covariance factor needs to be updated to make necessary adjustment for next iteration.

$$P_t = (I - K.H)P_{predicted} \quad (5)$$

2.3. Risk Assessment using Hierarchical Fuzzy Inference System (FIS)

Hierarchical FIS combines the conventional fuzzy systems in a layered(tree-like) structure. The whole system is made up of several individual fuzzy inference systems (FIS) which forms components of the whole system. Every FIS takes some inputs and generate some output. This several FIS are connected such that the output of bottom layer FIS serves as input for next layer FIS components. The connectivity structure is usually based on relationship among the parameters and overall structure needs not necessarily to be a tree because in some layers, one FIS component can feed input to more than one FIS in next layer thus violating the tree structure. This tree structure divides the whole problem into several pieces and thus reducing the total number of rules in the whole system.

The computation in hierarchical FIS is performed from bottom up and FIS at the top-level node provides the result for whole computation. Depending upon the problem nature, several configurations of FIS components can be designed, each forming a different model for the same problem. For this experiment, we consider a distinct model of hierarchical FIS as shown in Figure 3.

In this model, total eight inputs are used for four risk contributing factors i.e. two parameters for each factor (risk probability and risk severity). In conventional FIS system, all these inputs will be given to a single FIS in which (m+1)8 total number of rules will be required where m is the number of membership functions defined for each parameter.

This shows the number of rules in conventional FIS grows exponentially with increase in input parameters. However, in proposed model, these eights inputs are divided into four groups and each pair of input parameters is given as input to an FIS at next level. Afterwards, the output of these four FIS at bottom layer are given as input to an FIS at level-1 which finally generate the output result at level-0 i.e. final risk for any given location.

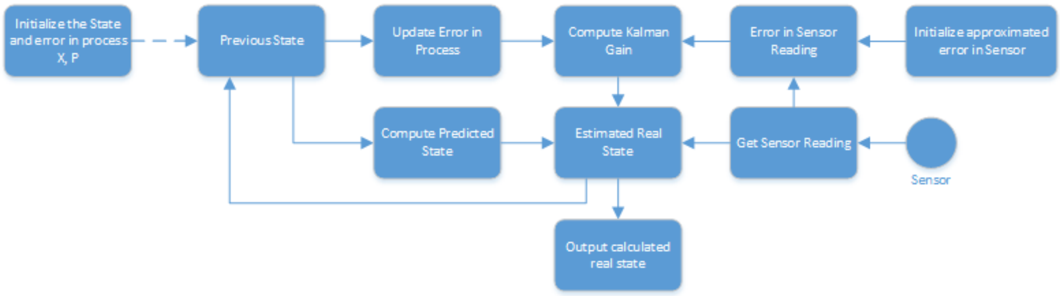

Figure 2. Kalman's Filter Algorithm.

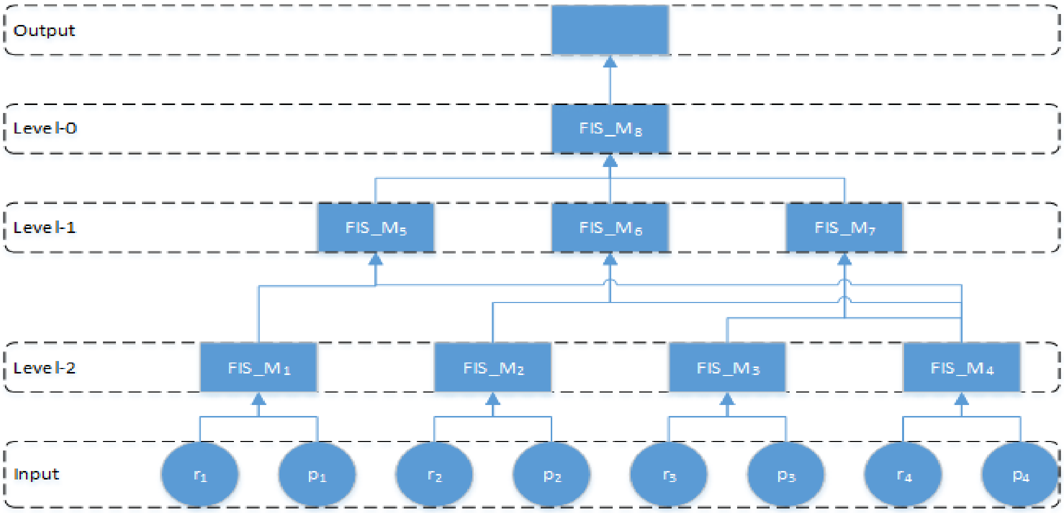

Figure 3. Integrated Hierarchical FIS.

In this model, we have used a unique tree structure i.e. M4 which is the ground state (geo-environmental) risk index is grouped with every other risk contributing factor at Level-1. Afterwards, the output of three FIS in Level-0 is combined to generate final risk index for any given location.

2.4. Configuration and Rules Definition for Hierarchical FIS

Rules definition is one of the most complex and important task in FIS based systems. It requires domain expert knowledge however, for automatic rules generation, we have used two different schemes. Both models are separately evaluated for sensor data, predicted data and original data.

Table 1: Membership functions labels and ranges for input and output parameters.

MF Name	MF Labels	Range
V. High	VH	(8.5 to 10)
High	H	(6.0 to 9.5)
Medium	M	(3.0 to 7.0)
Low	L	(0.5 to 4.0)
V. Low	VL	(0.0 to 1.5)

For every risk contributing factor, the risk values are scaled in [0, 10]. Five membership functions are defined within this range using triangular method and its details are given in Table 1. Membership functions are overlapping and equally distributed. Fuzzy membership function for the final output risk index are also the same. We have implemented a specific model in MATLAB to evaluate the performance of hierarchical FIS based model. Details of FIS components configurations are given in Table 2.

As we have two schemes for rules definition and therefore selected hierarchical FIS model is evaluated separately for each scheme. Following section presents brief illustration of both schemes.

Table 2: Configuration for various parts of individual FIS.

FIS Particulars	Definition of Membership function	Input Fuzzification	Implication Method	Aggregation Method	Defuzzification Method
Scheme Name	Membership functions are defined using Triangular	Min is used for AND operation and Max is used for OR operation	Min is used for result implication	Max is used for Aggregation method	Centroid method is used as defuzzification to compute final output

2.4.1. Rules definition for Hierarchical FIS

Usually, underground risk

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

assessment involves several associated risks contributing factors. We need a system that can somehow combine all related risk contributing factors to generate risk representative value for all. In this study, we have considered four risk contributing factors i.e. sewerage line risk index, water supply risk index, ground state risk index, and metro structure risk index. In Hierarchical FIS, these risk factors are processed in a layered fashion. Two or more risk factors are mixed using FIS to produce their effective risk index. Finally, from the root FIS, we get the resultant risk index for the desired location.

As discussed earlier that at the core of fuzzy systems, we have rules that control the whole process. Fuzzy rules in our hierarchical FIS, will somehow combine two or more risk index values to compute its corresponding single risk index value. For rules definition, two different approaches are adopted as below.

Average based: In this scheme, output value is set to the average value of given input e.g.

IF M_1 is low AND M_2 is high THEN y is Medium

Maximum based: In this scheme, output value is set to the maximum value of given input e.g.

IF M_1 is low AND M_2 is high THEN y is high

Table 3 presents summary of the fuzzy sub-systems in both hierarchical FIS models by expressing number of input to each fuzzy sub-system and its corresponding number of membership functions. Total number of rules for every fuzzy sub-system is counted by considering all possible combinations of input variables i.e. (Membership function) No. of variables. Here, we take power 6 (instead of 5) and this is to consider the null option for every variable. Finally, we subtract 1 and this is to ignore the combination where all inputs are null. Last column in Table 3 presents the total number rules in selected model.

Table 3: Rules setting for the selected Hierarchical FIS model.

Model #	FIS Sub System	Input Variables	Membership Functions	No. of Rules	Total No. of Rules
Integrated Hierarchical FIS	FIS_M_1	2	5	6^2-1=35	460
	FIS_M_2	2	5	6^2-1=35	
	FIS_M_3	2	5	6^2-1=35	
	FIS_M_4	2	5	6^2-1=35	
	FIS_M_5	2	5	6^2-1=35	
	FIS_M_6	2	5	6^2-1=35	
	FIS_M_7	2	5	6^2-1=35	
	FIS_M_8	3	5	6^3-1=215	

III. RESULTS AND DISCUSSION

3.1. Data Collection

For experimental purposes, we have synthetic data carefully generated for selected location over period of two months (60 days) in a controlled environment. Premeditated changes are made to surface and environmental condition to have maximum diversity in collected data for better analysis. We also have estimated final risk index for the selected location by experts for the given time. Risk index values are scaled in range [0, 10] where the minimum value 0express no risk and the highest value 10 represent very high risk. Estimated risk index values for the given location are shown in Figure 4.

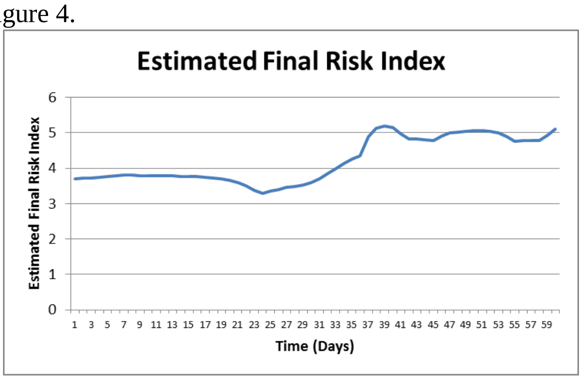

Figure 1. Estimated risk index for selected location over period of 02 months.

Furthermore, values of four risk contributing factors are estimated from sensing data which further serves as input data to our system i.e. M_1 for water supply risk index, M_2 for sewerage line risk index, M_3 for metro structure risk index, M_4 for ground state risk index. Sensor readings have noise which usually occurs because of hardware issues, imperfection in analysis algorithms, environmental conditions and battery power fluctuations further intensifies this problem. All these sources of error significantly damage the quality of collected data through sensors.

3.2. Risk Prediction Results using Kalman Filter Algorithm

Almost every data analysis process begins with data cleaning to prepare data for further processing. Data often includes noise, errors, mistakes, duplications, irrelevant information etc. During data cleaning, we try to remove erroneous and unnecessary data from the collected data. Any results based on uncleaned data are highly compromised and can lead to false or incorrect decision. Therefore, data cleaning is very crucial step in information processing and it includes noise removal, data transformation, handling missing values, redundant data elimination, etc. In modern IoT based systems, data is often collected via smart sensors. Sensing data can have errors due to various reasons e.g. poor calibration, power fluctuation, environmental conditions and errors due to data transmission in the network. Kalman filter algorithm is a very famous algorithm, commonly used to accurate estimation of true readings from noisy sensor data. Results of Kalman filter prediction for risk contributing factors is shown in Figure 5.

Published By: Blue Eyes Intelligence Engineering & Sciences Publication

Sensor data has too much variation (noise) which is effectively removed by Kalman filter algorithm. Kalman filter results are closer to actual original data.

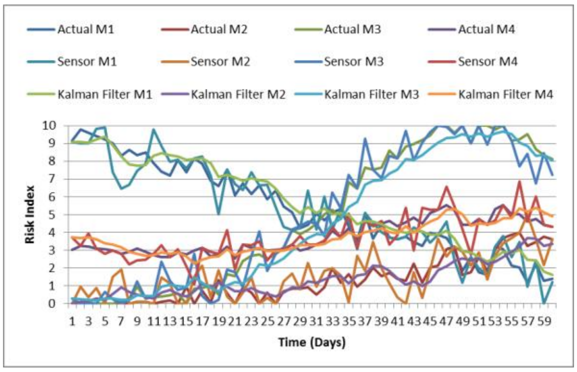

Figure 2. Risk index values using Kalman Filter for four selected risk contributing factors.

3.3. Results of Basic Hierarchical FIS

Figure 6 illustrates the output result for hierarchical model with real sensor data, data cleaned with Kalman filter and actual data. For all three kinds of input data, hierarchical FIS is evaluated twice using different scheme for FIS rules definition. average based and maximum based scheme. Results show that maximum based scheme is performing well as compared to average based scheme. Average based schemes fail to capture system desired results and it suppresses the final risk index. Maximum based HFIS scheme produces good results for all kinds of input data with exception to sensor data results whose final risk index has too much variations and fluctuations. Kalman filter results for maximum based HFIS schemes helps stabilizing the output risk index and its output is more approximate to result produced by actual data.

Results for average based scheme for rules definition are very poor as compared maximum-based scheme. This concludes that average based scheme does not work fine. This is since average based scheme suppresses the risk values at lower layers and hence the resultant final risk index is very low from the root level FIS. In contrast, maximum based scheme for rule definition gives much better results. It even outperforms the maximum based scheme results in model. This is since addition of extra layer in selected model where ground state risk factor is integrated individually with other three risk factors is much better representation of relationship among the risk factor. We can see that using sensor data without applying Kalman filter produces poor results, having frequent spikes and fluctuations. Kalman filter results for maximum based HFIS schemes helps stabilizing the output risk index and its output is more approximate to results produced by actual data and the three lines are nicely moving together with near perfect alignment. This proves that selected model with maximum based scheme produces best results with Kalman filter.

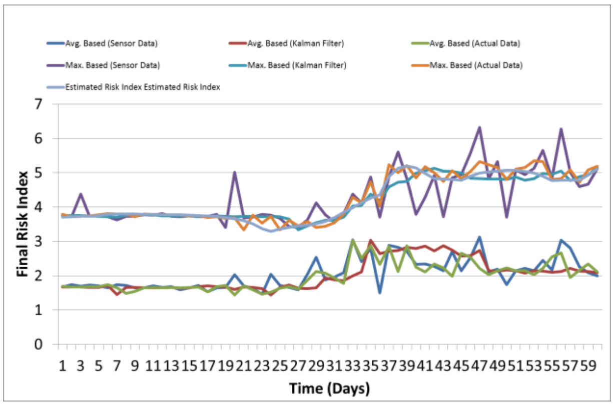

Figure 3. Output results of hierarchical basic model with average and maximum based schemes for rules definition using sensor data, actual data and Kalman filter predicted data.

3.5. Statistical Comparison

For statistical comparison collected results and analysis, we consider the following four measures along with their respective mathematical formula.

$$\text{Mean Absolute Deviation (MAD)} = \frac{\sum(x_i - \bar{x})}{n} \quad (6)$$

$$\text{Mean Square Error (MSE)} = \frac{\sum(x_i - \bar{x})^2}{n} \quad (7)$$

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} \quad (8)$$

$$\text{Pearson Correlation (R)} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (9)$$

We are calculating these measures for all three kinds of data (actual, sensor and Kalman filter) by comparing it with estimated risk index. Results are given in Figure 8.

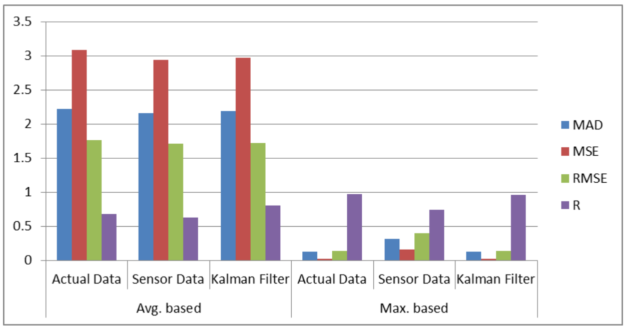

Figure 4. Statistical comparison final risk index with actual, sensing and Kalman filter data.

Figure 7 presents comparative analysis of the selected models with average based scheme and maximum based scheme in terms of four statistical measures. These results reveal that average based scheme for rules definition fails for selected model on all measures and considered out of the race. Maximum based scheme for rules definition seems to work fine in selected model. Comparatively, maximum based scheme for selected model outperform all other schemes having correlation of +0.960 with Kalman Filter and +0.968 with actual data.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

These results indicate that the layering structure in the selected model accurately captures the actual relationship among the individual risk contributing factors. Furthermore, among the two schemes for rules definition, maximum based scheme gives better results. Kalman filter helps in predicting near actual final risk index by removing the error (noise) in data. Furthermore, it can also be seen that for the same data, selected model results are better in all statistical measures which proves that selected model is more resilient and less effected by noisy data.

IV. CONCLUSION

In this paper, we have presented an improved model based on Kalman filter and hierarchical fuzzy inference system for risk prediction and assessment in underground facilities. For experimental analysis, we have used synthetic data carefully generated for 60 locations by considering four risk contributing factors i.e. sewerage line risk index, water supply risk index, ground state risk index, and metro structure risk index. A unique model based on hierarchical FIS is evaluated in this study using two different schemes for rules generation i.e. average based scheme and maximum based scheme. Results indicate that the selected model with maximum based scheme for fuzzy inference system rules definition outperform the other scheme. Furthermore, significant improvement in risk prediction results is observed with Kalman filter scheme. Execution time of the selected model is also significantly lower than the conventional FIS due to reduced number of rules. In future, we will use real data to verify the proposed model for underground risk assessment and prediction.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2018R1D1A1A09082919), and this work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT)(No.2018-0-01456, AutoMaTa: Autonomous Management framework based on artificial intelligent Technology for adaptive and disposable IoT). Any correspondence related to this paper should be addressed to DoHyeun Kim.

REFERENCES

1. Carpenter M, Erdogan B, Bauer T. Principles of Management. Flat World Knowledge. Inc. USA. 2009 Dec 5.
2. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism: prospective cohort study. BMJ. 2011 Aug 16;343:d4656.
3. Saaty TL. Decision making with the analytic hierarchy process. International journal of services sciences. 2008 Jan 1;1(1):83-98.
4. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology (Cambridge, Mass.). 2010 Jan;21(1):128.
5. Keller JM, Gray MR, Givens JA. A fuzzy k-nearest neighbor algorithm. IEEE transactions on systems, man, and cybernetics. 1985 Jul(4):580-5.
6. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intelligent Systems and their applications. 1998 Jul;13(4):18-28.
7. Breiman L. Random forests. Machine learning. 2001 Oct 1;45(1):5-32.
8. Hagan MT, Demuth HB, Beale MH, De Jesús O. Neural network design. Boston: Pws Pub.; 1996.
9. Harvey AC. Forecasting, structural time series models and the Kalman filter. Cambridge university press; 1990.
10. Delmastro C, Lavagno E, Schranz L. Underground urbanism: master plans and sectorial plans. Tunnelling and Underground Space Technology. 2016 May 1;55:103-11.
11. Broere W. Urban underground space: Solving the problems of today's cities. Tunnelling and Underground Space Technology. 2016 May 1;55:245-8.
12. Nallapaneni Manoj Kumar, Sonali Goel, Pradeep Kumar Mallick, "Smart Cities in India: Features, Policies, Current Status, and Challenges" , IEEE International Conference on Technologies for Smart-City Energy Security and Power (ICSESP 2018), C. V. Raman College of Engineering, Bhubaneswar, March 28-30, 2018., Electronic ISBN: 978-1-5386-4581-9, DOI: 10.1109/ICSESP.2018.8376669.
13. Ullah I, Fayaz M, Kim D. Analytical modeling for underground risk assessment in smart cities. Applied Sciences. 2018 Jun;8(6):921.