Journal of Electronics and Informatics (2021)
Vol.03/ No.02
Pages: 126-137
https://www.irojournals.com/iroei/
DOI: https://doi.org/10.36548/jei.2021.2.005

Industrial Quality Prediction System through Data Mining Algorithm

P. Karthigaikumar,
Professor,
Department of Electronics and Communication Engineering,
Karpagam College of Engineering,
Coimbatore, India.

Abstract: Based on an assessment of production capabilities, manufacturing sectors' core competency is increased. The importance of product quality in this aspect cannot be overstated. Several academics have introduced Deming's 14 principles, Shewhart cycle, total quality management, and other approaches to decrease the external failure costs and enhance product yield rates. Analysis of industrial data and process monitoring is becoming increasingly important as a part of the Industry 4.0 paradigm. In order to reduce the internal failure cost and inspection overhead, quality control (QC) schemes are utilized by industries. The final product quality has an interactive and cumulative effect of various parameters like operators and equipment in multistage manufacturing processes (MMP). In other cases, the final product is inspected in a single workstation with QC. It's challenging to do a cause analysis in MMP whenever a failure occurs. Several industries are looking for the optimal quality prediction model in order to achieve flawless production. The majority of current approaches solely handles single-stage manufacturing and is inadequate in dealing with MMP quality concerns. To overcome this issue, this paper proposes an industrial quality prediction system with a combination of multiple Program Component Analysis (PCA) and Decision Stump (DS) algorithm for MMP quality prediction. A SECOM (SEmiCOnductor Manufacturing) dataset is used for verification and validation of the proposed model. Based on the findings, it is clear that this model is capable of performing accurate classification and prediction in the field of industrial quality.

Keywords: Quality control, principal components analysis, decision stump, multistage manufacturing process, data mining, quality prediction

ISSN: 2582-3825 (online)
Submitted: 21.05.2021
Revised: 13.06.2021
Accepted: 29.06.2021
Published: 08.07.2021

Journal of Electronics and Informatics (2021)
Vol.03/ No.02
Pages: 126-137
https://www.irojournals.com/iroei/
DOI: https://doi.org/10.36548/jei.2021.2.005

1. Introduction

Based on the manufacturing operation condition, the quality of the finished product must be predicted by using an on-line quality monitoring model [1]. Quality prediction can be performed by enabling a mathematical model or formulation that relates to the industrial manufacturing operation condition. This system is termed as a quality prediction framework [2]. Evaluation of the manufacturing operation is performed by monitoring the product quality level by process engineers with the help of a quality prediction model. Historical manufacturing datasets are used for developing quality prediction models with the help of several data mining techniques in the recent years. Various industries make use of regression, association rules, classification and clustering schemes. Food processing industry, loudspeaker manufacturing, hard disk manufacturing, machine process, slider manufacturing, semiconductor manufacturing, and injection molding industries have implemented these techniques [3]. The requirements of Single-stage Manufacturing Processes (SMP) are addressed by using these prediction models.

The real-world industrial scenario is driven by multi-stage manufacturing processes (MMP) in the recent years [4-5]. In order to manufacture the complex products, multiple workstations are involved in the industrial manufacturing process. This framework is termed as MMP. The popularity of MMP is growing as there is a continuous growth in the product structure complexity along with the sophistication in the taste of client [6]. With regard to the complex structure, multiple stages of manufacturing is involved in the production of aerospace devices, automotive products, semiconductor, and printed circuit board (PCB). In MMP, quality prediction models are developed to attain a faultless manufacturing [7-9]. The SMP approach is used for the development of MMP approach in several quality prediction models. Multiple workstations perform a series of manufacturing operations in order to produce a final product in MMP. The cumulative effect of workstations may lead to ineffective and misleading quality measurement in MMP while using SMS approach. This is due to the effect of previous workstations and the preceding manufacturing operation that exist in the next workstation. Quality prediction is also performed by using the Cascade Quality Prediction

ISSN: 2582-3825 (online)
Submitted: 21.05.2021
Revised: 13.06.2021
Accepted: 29.06.2021
Published: 08.07.2021

Journal of Electronics and Informatics (2021)
Vol.03/ No.02
Pages: 126-137
https://www.irojournals.com/iroei/
DOI: https://doi.org/10.36548/jei.2021.2.005

Method (CQPM) framework in certain scenarios [10]. However, the model and its accuracy have not been thoroughly investigated.

2. Related Works

Two alternative techniques are used for MMP based quality prediction model development. For the entire manufacturing line, a single prediction model has been developed as the first alternative [11]. Manufacturing operation that occurs in every workstation is considered as if it is occurred in a single workstation by using the single-point approach. Association rules, clustering, classification and several other data mining techniques are employed for improving the MMP based quality prediction architecture [12]. The second approach is based on the development of customized prediction models for individual workstations. It is also termed as a multi-point approach. The entire manufacturing line makes use of multiple prediction models by using this approach. Partial Least Square (PLS), Principal Component Analysis (PCA), clustering and other techniques are employed in this approach for the development of a prediction model [13-15]. The quality level of the product is assumed to be independent at each manufacturing workstation while applying data mining schemes or multivariate statistics by using the single-point approach. Using this approach, determining the relationship between workstations and manufacturing activities is difficult. From the perspective of total quality, this method partially explains the quality of the last workstation [16].

A particular workstation and its behavior may be modeled by using the multi-point approach. The model developed by this technique may explain the workstation variables and their relationships throughout manufacturing operations [17]. The previous workstations and their cumulative effect is confounded with the measurements of a particular workstation leading to ineffective and misleading results. For a specific workstation, the partial quality may be explained by the individual stage models. Approaches other than multi-point and single-point are also proposed by certain researchers. MMP based quality prediction is performed by developing a CQPM based cascade approach [18]. Product characteristics are used for representing the final product quality using this approach. The MMP based variable

ISSN: 2582-3825 (online)
Submitted: 21.05.2021
Revised: 13.06.2021
Accepted: 29.06.2021
Published: 08.07.2021

Journal of Electronics and Informatics (2021)
Vol.03/ No.02
Pages: 126-137
https://www.irojournals.com/iroei/
DOI: https://doi.org/10.36548/jei.2021.2.005

relationships form the major element of the design in this model. The MMP based variable relationship conditions are explained under certain basic criteria [19]. The output quality from specific workstation is influenced by the workstations manufacturing operation variables. The previous workstation outputs as well as the manufacturing process affects the workstations output quality. The operation variables of manufacturing influences the final product quality. Random Forest (RF) [20], Random Tree (RT) [21], Chi-squared Automatic Interaction Decision (CHAID) [22], Decision Stump (DS) [23], C4.5 [24], and Iterative Dichotomiser 3 (ID3) [25] are some of the decision tree algorithms that are available. The dataset is used for building decision tree using supervised learning classification algorithms. The classes are mapped with the attribute values using diagrams built by these algorithms based on a set of pre-classified cases [26]. Different splitting techniques and split criteria are used by the six decision tree algorithms despite operating it in a similar manner.

3. Methodology

LISP evaluation and classification algorithms are implemented in this work and compared with the existing models for performing quality prediction and analysis. Both single point and multi point approaches are discussed while considering the MMP characteristics. The training and testing datasets are categorized from the original manufacturing dataset after the noise removal and pre-processing stage. Further, the LISP mining and classification process is used for building a quality prediction model. The MMP property is represented in an efficient manner by performing appropriate selection of feature set and features so as the relationship between the impact on the final product and the workstations are analyzed by using classifier learning. The confidence threshold and minimum support are defined on the training dataset and then, the apriori algorithm is applied for LISP mining. The testing dataset is used for the evaluation of association rules and output classifiers. In multiple workstation scenario, the defect cause analysis and quality prediction is performed online by using this model. The proposed methodology design flow is represented in figure 1.

ISSN: 2582-3825 (online)
Submitted: 21.05.2021
Revised: 13.06.2021
Accepted: 29.06.2021
Published: 08.07.2021

Journal of Electronics and Informatics (2021)
Vol.03/ No.02
Pages: 126-137
https://www.irojournals.com/iroei/
DOI: https://doi.org/10.36548/jei.2021.2.005

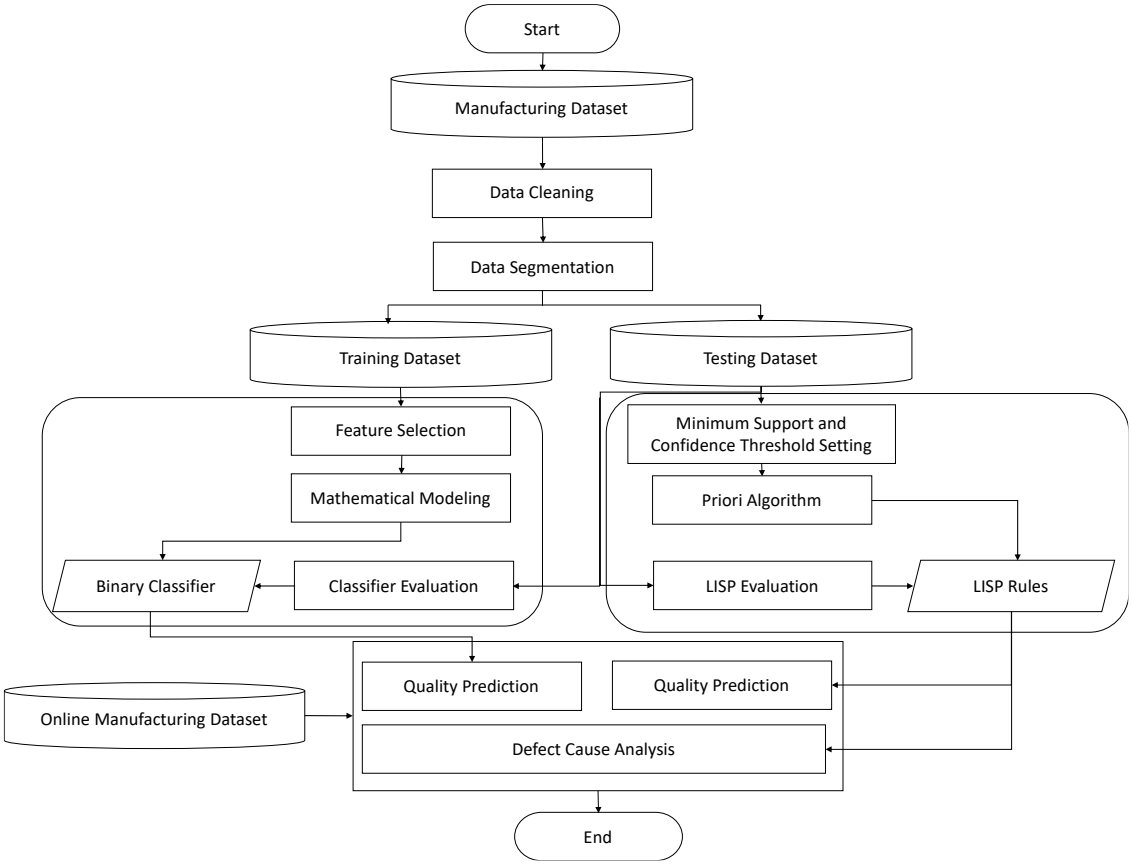

Fig. 1. Proposed Quality Prediction Model Flow

It is essential to include normalization and cleaning in data-preprocessing stage as several situations and events consist of various anomaly in the manufacturing environment. From all instances, mean value is calculated that replaces the missing values to overcome the issue of missing value. If discriminative information does not exist in a variable, it will be eliminated when similar instance values are observed in variables from all instances. The LISP rules between independent and target variables are mined using dependency modeling. The LISP rule mining refers to an active financial data mining technique. Priori algorithm is a classical algorithm used in this approach. Here, aggressive building of relations is performed when frequent items are identified. The item sets that are less frequent are removed and the rules are pruned by the algorithm. Physical facts are used as the basis in certain algorithms. Temperature and power parameters of IC chips and their relationship is described using

ISSN: 2582-3825 (online)
Submitted: 21.05.2021
Revised: 13.06.2021
Accepted: 29.06.2021
Published: 08.07.2021

Journal of Electronics and Informatics (2021)
Vol.03/ No.02
Pages: 126-137
https://www.irojournals.com/iroei/
DOI: https://doi.org/10.36548/jei.2021.2.005

physics based model built based on dependency modeling. Partial least square analysis and principle component analysis contribution plot are some of the other approaches used.

Failed products and insufficient instances for them is another significant issue in research related to quality prediction. A high yield rate is obtained for a stable and typical shop floor in the recent manufacturing lines. Boosting techniques and oversampling is introduced for balancing the positive and negative dataset sizes. The fail dataset is simply duplicated to similar amount of pass dataset while adopting the boosting scheme. The inter-correlated variables relationships are identified based on the definition of CQPM. The MMP characteristics are carried while transforming the latent variables into a new dimension set using the PCA. The binary classifier is trained by applying Naïve Bayes, support vector machine (SVM), decision tree and other classification algorithms after selection of appropriate feature set. For decision tree learning, potential algorithms like C50, recursive partitioning and regression trees (RPART) and iterative dichotomiser 3 (ID3) are used.

4. Results and Discussion

SECOM (SEmiCOnductor Manufacturing) dataset is used for verifying the proposed framework. A single quality variable and 590 manufacturing operation variables are present in each of the 1567 samples present in the dataset. Failure case is represented by 104 samples present in the entire dataset. Based on sensor ID, the sensors are named in semiconductor manufacturing line. These process control sensors collect variable data from each operation. While applying boosting to the dataset in order to balance the negative and positive dataset, 1456 unsuccessful occurrences were discovered. On a 1:3 basis, from the original dataset, segmentation of the testing and training dataset is done. At the data pre-processing level, 115 redundant features are removed initially for pre-processing the data. For classifier learning, the most significant 40 features are retained after the feature extraction stage.

The semiconductor monitoring process property is used for categorizing the 40 features into five groups for MMP scenario simulation while each group represents a workstation. Features with correlation is obtained by applying the dataset with PCA prior to classification. The

ISSN: 2582-3825 (online)
Submitted: 21.05.2021
Revised: 13.06.2021
Accepted: 29.06.2021
Published: 08.07.2021

Journal of Electronics and Informatics (2021)
Vol.03/ No.02
Pages: 126-137
https://www.irojournals.com/iroei/
DOI: https://doi.org/10.36548/jei.2021.2.005

principle component (PC) is obtained through PCA by combining the features with previous workstation at every workstation. A single quality result label and 14 correlation features are obtained in the final dataset. The SECOM dataset is verified with six models as represented in figure 2. The proposed model is compared with the existing state-of-the-art models. Accuracy, Gmean, True positive rate, false alarm rate, F1 score and balanced error rate are considered to be the evaluation metrics and compared for performance. The confusion matrix for the dataset is estimated based on the expressions for these metrics as represented by expression (1) to (6).

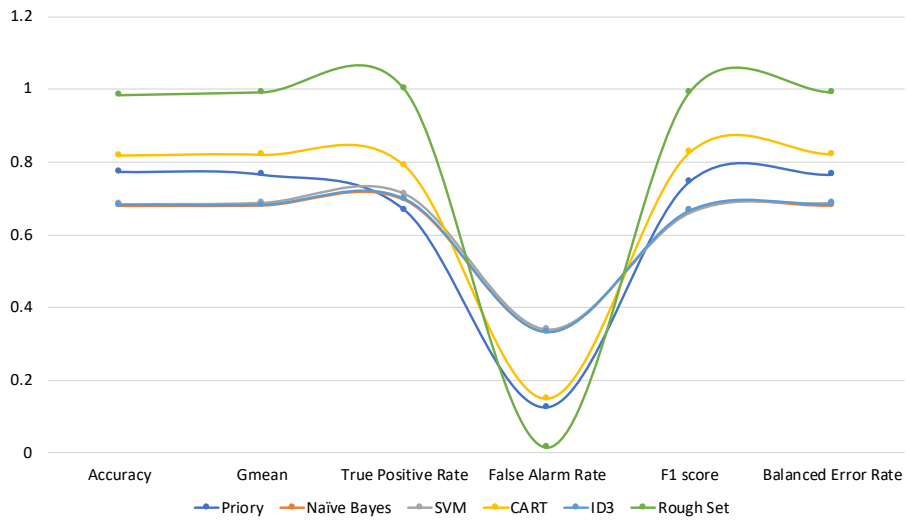

Fig. 2. Comparison of parameters of the existing and proposed model

$$\text{Accuracy} = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \text{ ------ (1)}$$

$$G_{mean} = \sqrt{\frac{True\ Positive}{True\ Positive + False\ Negative} \times \frac{True\ Negative}{True\ Negative + False\ Positive}} \text{ ------- (2)}$$

$$\text{True Positive Rate} = \frac{True\ Positive}{True\ Positive + False\ Negative} \text{ ------ (3)}$$

ISSN: 2582-3825 (online)
Submitted: 21.05.2021
Revised: 13.06.2021
Accepted: 29.06.2021
Published: 08.07.2021

Journal of Electronics and Informatics (2021)
Vol.03/ No.02
Pages: 126-137
https://www.irojournals.com/iroei/
DOI: https://doi.org/10.36548/jei.2021.2.005

$$\text{False Alarm Rate} = \frac{False\ Positive}{False\ Positive + True\ Negative} \text{ ------ (4)}$$

$$\text{F1 score} = \frac{2 \times True\ Positive}{(2 \times True\ Positive + False\ Positive + False\ Negative)} \text{ -------- (5)}$$

$$\text{Balanced Error Rate} = 1\text{-}0.5 \times \left(\left(\frac{False\ Negative}{True\ Positive + False\ Negative} \right) + \left(\frac{False\ Positive}{True\ Negative + False\ Positive} \right) \right) \text{ ----(6)}$$

Figure 3 represents the comparison of the decision tree algorithms performance. It is observed that CART and the proposed algorithm offers highest true positive rate and least false negative rate. These values are used for estimation of other performance parameters. During manufacturing stage, it is possible to detect all failed products with a true positive rate of 1. When compared to the existing models, the proposed rough set model offers have improved performance among all parameters. The minimum support threshold and confidence are passed with the LISP for building the quality prediction model. More than one rule is met by several instances. Further, accurate prediction performance may be obtained by improving the involvement of knowledge expertise while understanding the variables physical meaning.

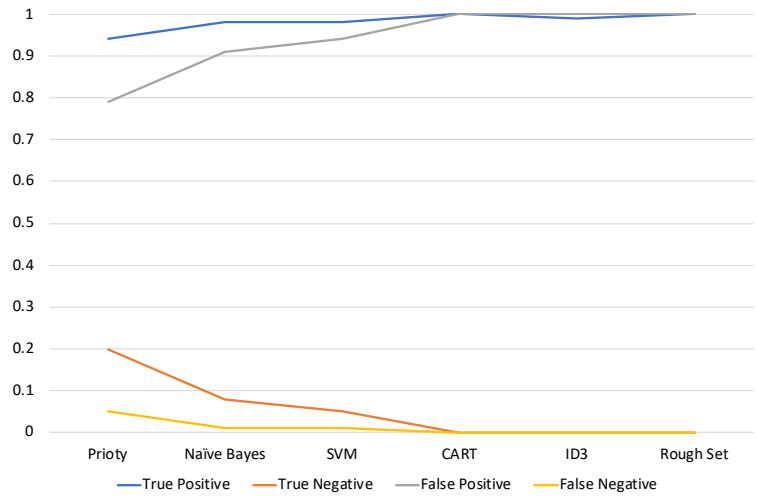

Fig. 3. Decision Tree Algorithms Performance comparison

ISSN: 2582-3825 (online)
Submitted: 21.05.2021
Revised: 13.06.2021
Accepted: 29.06.2021
Published: 08.07.2021

Journal of Electronics and Informatics (2021)
Vol.03/ No.02
Pages: 126-137
https://www.irojournals.com/iroei/
DOI: https://doi.org/10.36548/jei.2021.2.005

5. Conclusion

For semiconductor manufacturing process, this research work has developed a quality prediction model by using single-point prediction scheme as well as CQPM for comparison. The G_{mean} value is highest for CQPM model based on the comparison results. While classifying the minority and majority classes, when compared to the single-point method, better prediction model is obtained by using the CQPM scheme. Relationship between the quality level of the final product and characteristic variables of the product are revealed by using the decision tree algorithms. The total values range from very low to very high in a uniform manner while using decision tree algorithm for comparing the attributes of the characteristic variables of all products. While classifying minority classes, RF, RT, DS, CHAID and C4.5 were a failure, while ID3 offered better performance with uniform attribute value count and imbalanced datasets. An improved prediction model by combining modified multiple PCA and ID is presented in this paper. In negative class, misclassification probability is high with relatively low G_{mean} and high accuracy values. Further, the performance of the model may be improved on a technical level. While encountering the imbalanced datasets, the performance of prediction model can be further improved in the future while combining additional techniques.

References

[1] Thiede, S., Turetskyy, A., Kwade, A., Kara, S., & Herrmann, C. (2019). Data mining in battery production chains towards multi-criterial quality prediction. CIRP Annals, 68(1), 463-466.

[2] Suresh, PP Jashma, U. Dinesh Acharya, and NV Subba Reddy. "Study of Effective Mining Algorithms for Frequent Itemsets." Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020 57 (2021): 499.

[3] Adam, Edriss Eisa Babikir. "Evaluation of Fingerprint Liveness Detection by Machine Learning Approach-A Systematic View." Journal of ISMAC 3, no. 01 (2021): 16-30.

ISSN: 2582-3825 (online)
Submitted: 21.05.2021
Revised: 13.06.2021
Accepted: 29.06.2021
Published: 08.07.2021

Journal of Electronics and Informatics (2021)
Vol.03/ No.02
Pages: 126-137
https://www.irojournals.com/iroei/
DOI: https://doi.org/10.36548/jei.2021.2.005

[4] Syafrudin, M., Fitriyani, N. L., Li, D., Alfian, G., Rhee, J., & Kang, Y. S. (2017). An open source-based real-time data processing architecture framework for manufacturing sustainability. Sustainability, 9(11), 2139.

[5] Bai, Y., Sun, Z., Zeng, B., Long, J., Li, L., de Oliveira, J. V., & Li, C. (2019). A comparison of dimension reduction techniques for support vector machine modeling of multi-parameter manufacturing quality prediction. Journal of Intelligent Manufacturing, 30(5), 2245-2256.

[6] Meenaatchi, S. M., and K. Rajeswari. "Rational Against Irrational Causes of Symptom Recognition Using Data Taxonomy." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 583-591. Springer Singapore, 2021.

[7] Shang, C., & You, F. (2019). Data analytics and machine learning for smart process manufacturing: recent advances and perspectives in the big data era. Engineering, 5(6), 1010-1016.

[8] Smys, S., and Jennifer S. Raj. "Analysis of Deep Learning Techniques for Early Detection of Depression on Social Media Network-A Comparative Study." Journal of trends in Computer Science and Smart technology (TCSST) 3, no. 01 (2021): 24-39.

[9] Sirisha Devi, J., and P. Vijaya Bhaskar Reddy. "Multimodal Emotion Analytics for E-Learning." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 593-602. Springer Singapore, 2021.

[10] Joe, Mr C. Vijesh, and Jennifer S. Raj. "Location-based Orientation Context Dependent Recommender System for Users." Journal of trends in Computer Science and Smart technology (TCSST) 3, no. 01 (2021): 14-23.

[11] Lingitz, L., Gallina, V., Ansari, F., Gyulai, D., Pfeiffer, A., Sihn, W., & Monostori, L. (2018). Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer. Procedia Cirp, 72, 1051-1056.

[12] Cheng, Y., Chen, K., Sun, H., Zhang, Y., & Tao, F. (2018). Data and knowledge mining with big data towards smart production. Journal of Industrial Information Integration, 9, 1-13.

[13] Vijayakumar, T., Mr R. Vinothkanna, and M. Duraipandian. "Fusion based Feature Extraction Analysis of ECG Signal Interpretation–A Systematic Approach." Journal of Artificial Intelligence 3, no. 01 (2021): 1-16.

ISSN: 2582-3825 (online)
Submitted: 21.05.2021
Revised: 13.06.2021
Accepted: 29.06.2021
Published: 08.07.2021

Journal of Electronics and Informatics (2021)
Vol.03/ No.02
Pages: 126-137
https://www.irojournals.com/iroei/
DOI: https://doi.org/10.36548/jei.2021.2.005

[14] Amara, Indraneel, K. Sai Pranav, and H. R. Mamatha. "Hybrid Recommendation System for Scientific Literature." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 725-735. Springer Singapore, 2021.

[15] Haoxiang, Wang, and S. Smys. "Big Data Analysis and Perturbation using Data Mining Algorithm." Journal of Soft Computing Paradigm (JSCP) 3, no. 01 (2021): 19-28.

[16] Lee, J. Y., Yoon, J. S., & Kim, B. H. (2017). A big data analytics platform for smart factories in small and medium-sized manufacturing enterprises: An empirical case study of a die casting factory. International Journal of Precision Engineering and Manufacturing, 18(10), 1353-1361.

[17] Vijayakumar, V. A., J. Shanthini, and S. Karthick. "Convolutional Recurrent Neural Network Framework for Autonomous Driving Behavioral Model." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 761-772. Springer Singapore, 2021.

[18] Chua, Z. Y., Ahn, I. H., & Moon, S. K. (2017). Process monitoring and inspection systems in metal additive manufacturing: Status and applications. International Journal of Precision Engineering and Manufacturing-Green Technology, 4(2), 235-245.

[19] Chen, Joy Iong Zong, and P. Hengjinda. "Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method-A Comparative Study." Journal of Artificial Intelligence 3, no. 01 (2021): 17-33.

[20] Zhang, Y., Ma, S., Yang, H., Lv, J., & Liu, Y. (2018). A big data driven analytical framework for energy-intensive manufacturing industries. Journal of Cleaner Production, 197, 57-72.

[21] Smys, S., Abul Bashar, and Wang Haoxiang. "Taxonomy Classification and Comparison of Routing Protocol Based on Energy Efficient Rate." Journal of ISMAC 3, no. 02 (2021): 96-110.

[22] Lade, P., Ghosh, R., & Srinivasan, S. (2017). Manufacturing analytics and industrial internet of things. IEEE Intelligent Systems, 32(3), 74-79.

[23] Shakya, Subarna. "Process mining error detection for securing the IoT system." Journal of ISMAC 2, no. 03 (2020): 147-153.

ISSN: 2582-3825 (online)
Submitted: 21.05.2021
Revised: 13.06.2021
Accepted: 29.06.2021
Published: 08.07.2021

Journal of Electronics and Informatics (2021)
Vol.03/ No.02
Pages: 126-137
https://www.irojournals.com/iroei/
DOI: https://doi.org/10.36548/jei.2021.2.005

[24] Ranganathan, G. "Real time anomaly detection techniques using pyspark frame work." Journal of Artificial Intelligence 2, no. 01 (2020): 20-30.

[25] Shakya, Subarna, Lalitpur Nepal Pulchowk, and S. Smys. "Anomalies Detection in Fog Computing Architectures Using Deep Learning." Journal: Journal of Trends in Computer Science and Smart Technology March 2020, no. 1 (2020): 46-55.

[26] Chung, K., Yoo, H., Choe, D., & Jung, H. (2019). Blockchain network based topic mining process for cognitive manufacturing. Wireless Personal Communications, 105(2), 583-597.

Author's Biography

P. Karthigaikumar has a rich experience of about 15 years in both teaching and academic administration. He has published 72 research papers in various International Journal and Conferences and also filed two patents. His areas of interest include VLSI Design, Control Systems etc., He was awarded ‘Best Faculty Award’ and ‘K.S. Krishnan Memorial Award’ from IETE in the year 2010.

ISSN: 2582-3825 (online)
Submitted: 21.05.2021
Revised: 13.06.2021
Accepted: 29.06.2021
Published: 08.07.2021