Journal of Green Engineering (JGE)

Volume-10, Issue-12, December 2020

# Process Analysis on Large Scale Manufacturing Industry for Performance and Sustainable Development

[1]Soumitra Singh, [2]Partha Sarathi Chakraborty, [3]S. Nallusamy, [4]K. Balakannan

*[1]Research Scholar, Department of Adult, Continuing Education & Extension, Jadavpur University, Kolkata, West Bengal, India.*
*E-mail: soumitra.singha@gmail.com*
*[2]Professor, Department of Adult, Continuing Education & Extension, Jadavpur University, Kolkata, West Bengal, India.*
*E-mail: p_s_c2001@yahoo.com*
*[3]Professor, Department of Mechanical Engineering, Dr. M.G.R. Educational and Research Institute, Chennai, Tamilnadu, India.*
*E-mail: ksnallu@gmail.com*
*[4]Principal, Adhiparasakthi College of Engineering, Vellore, Tamilnadu, India.*
*E-mail: kbalakannan@gmail.com*

## Abstract

Process mining is used to promote the business and production in organization through raw data pre-processing, relevant data usage by feature selections, mining methods, technique, simulations, validation and testing and sustainable manufacturing. Process analysis is vital role to convert prototype into product. Main problem in manufacturing industries are migrating between business and IT perspective and hard to automate with green manufacturing. Also achieving computerized clarifications from high-level production representation and analyzing the implementation of processes from a technical data and a business perception is widely difficult task. Continues evaluation of process mining and performance monitoring in components levels are very essential benchmark in large scale industries. Data mining is one of the recent approaches to improve the performance of the output and business. This research promotes the business and product output with sustainable manufacturing of the car manufacturing industries

*Journal of Green Engineering, Vol. 10_12, 12737-12752.*

with the help of machine learning algorithm. It mined the information about motor vehicles and also controls the fuel usage with green engineering by proposed model and the manufactures may select the model of the car based on the customer usage point through clustering concept.

**Keywords:** Green Manufacturing, Data Mining, Machine Learning, Process Analysis, Classification and WEKA

## 1 Introduction

This research work focuses on applied data mining algorithm for large scale industry and integrates the business intelligence to production process, promote the product and increase the world economic market.

### *1.1 Process Mining and Analysis*

Process mining facilitates to examine the business progression automatically based on the episode logs from Customer Relationship Management (CRM) and Enterprise Resource Planning (ERP) organizational data. It produces useful information from hidden patterns and visual mapping of organization activities to promote feature orders. Process analysis is vital role in production stage of different phase that revolves input into output to increase production effectively in any organization. It is an assessment to promote an organization to the top level. It reduces wastage of materials, labour time and avoids unnecessary activities. Figure 1 shows the process steps from raw material levels to production levels in manufacturing industries.

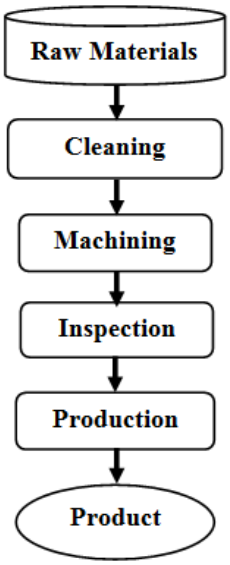


**Figure 1** Component Manufacturing in Manufacturing Organization

### *1.2 Data Mining*

Data mining discovers hidden patterns from raw data of manufacturing industry and forecast future outcomes with sustainable development based on earlier unidentified associations. Functionalities of data mining are clustering, classification, feature selection and outer layer analysis. Clustering could be done by statistical domain utilized for control and signal processing, information theory and mining technologies. According to perspective of Machine Learning (ML) the cluster shows the hidden pattern through the unsupervised learning approach. For realistic point of view clustering participates imperative in data mining applications such as industry big data exploration, fast information retrieval by semantic, cybernetics, ERP and CRM data in large scale industry.

#### *1.2.1 Activities in Data Mining*

- Data cleaning stage, raw data noises are removed by using soft approaches. Irrelevant attributes are removed in the appropriate stage.
- Relevance analysis stage, correlation measures are used to identify the relations between attributes.
- Data transformation and reduction stage, normalization and generalization are applied to create concept hierarchies.

### *1.3 Objective of the Research*

Objective of this research is to develop the machine learning approaches to control the fuel usage and manufacture can select the car models based on customer interests. Due to this, identifying bottleneck, redundancies and removing unwanted resources with help of machine learning algorithms. It performs as a tool to preserve and progress the industry processes and facilitate to achieve the incremental to revolution promotions, cost less, proper resource, valuable human resource management, sustainable development and increase of production throughput.

## 2 Literature Review

The organization improving production scheduling and organize activities, can guide to a worldwide expansion of industrialized schemes. For handling large storage capacity data availability and requirement of computing power, the machine learning techniques is an attractive elucidation to undertake manufacturing challenges [1-6]. Process manufacturing is take part in supporting the development of the international economy and preserving social profits [7-9]. At the upper stage of preparation and scheduling, taking decisions by ML approaches are goes to

improve economic profits even though raising global contest. Zhiqiangge et al developed ML model to analyses the organizational data [10]. The ML approach of supervised learning, unsupervised learning and also semi supervised learning applied for manufacturing environment [11-12]. Supervised learning is used for density estimation, dimensionality reduction, data visualization and process control. Unsupervised learning is applied for fault classification, quality estimation and key performance index prediction. The regression data can analysis with help of semi supervised algorithm. Naive Bayes classifier is based on Bayes' theorem, contain group of classification algorithms, can implemented effectively.

Applied clustering concept is used to analysis of customer patterns and wine dataset with help of WEKA tool [13-14]. Danyang et al [15] presented vehicle daily mobility patterns by analysis of car data. For classifying the vehicles characteristics and types, analysis is carried out to identify the exceptional features of clustered group. The distributed data mining is used to promote vehicle telematics products [16, 17]. Kingsley et al [18] identified a new semantic approach previous activity in the form of query to improve the process task. Kramberger [19] developed deep learning model using Stanford car dataset to increase the performance in industry.

## 3 Materials and Methods

This research uses as benchmark for car dataset analysis using data mining and machine learning algorithms.[20,21] Proposed model is used for clustering and classification data mining approach. Relevant features are clustered by K-means and vehicles fuel usage is identified by ZeroR and IBK classification algorithms [22].

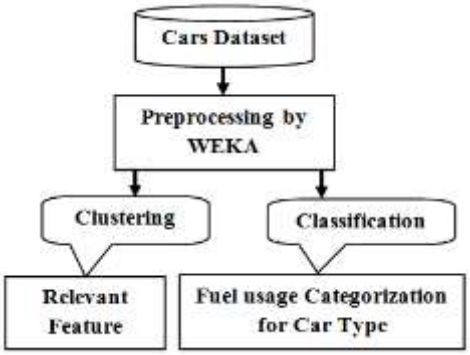


**Figure 2** Proposed Model

Figure 2 represents the proposed model architecture. Data set was taken from public website as https://corgis-edu.github.io/corgis/csv/cars/.y. It contains 5076 instances and 18 features as shown Figure 3 and Figure 4

respectively. It works for unsupervised and supervised tasks. Proposed model is implemented using WEKA data mining tool.

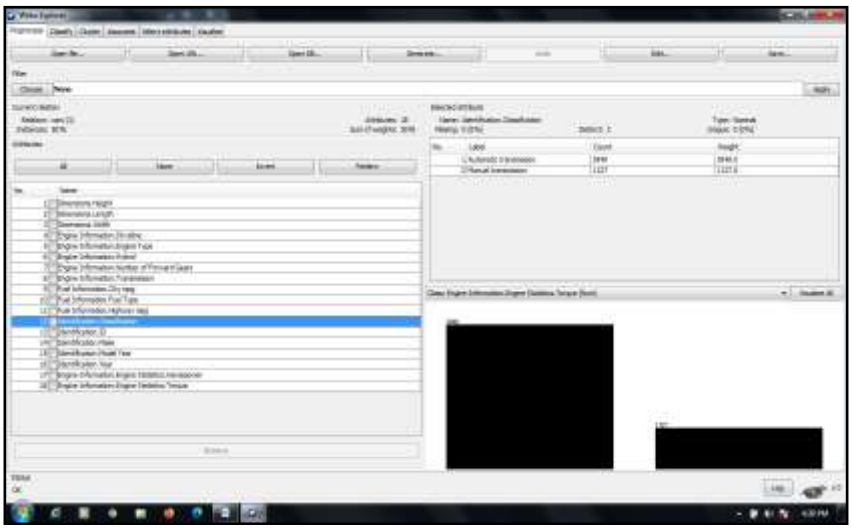

**Figure 3** Car.CSV

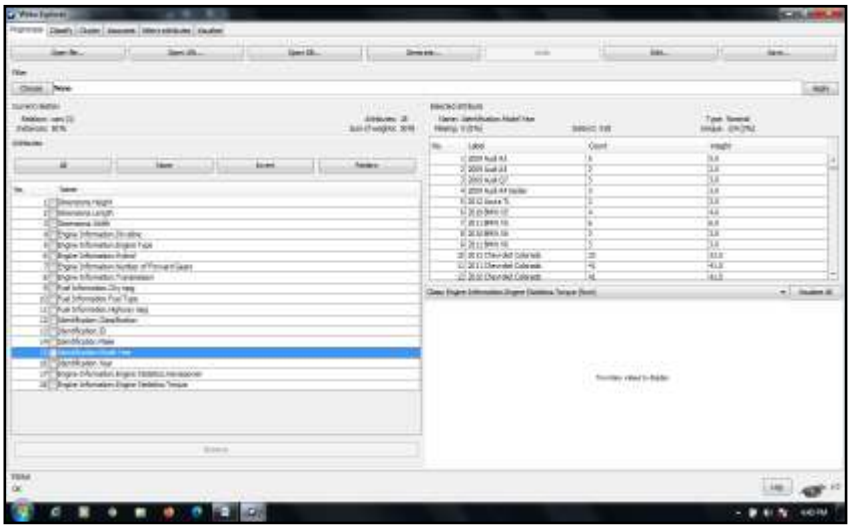

**Figure 4** Features of Cars Data

WEKA is an open source tool and it contains enormous machine learning algorithms. It accepts only Comma Separated Values (CSV) files and Attributes File Format (ARFF) files. Table 1 shows the some of the example attributes and data in ARFF.

**Table 1** ARFF of Cars with CSV

relation 'cars'
@attribute Dimensions.Height numeric
@attribute Dimensions.Length numeric
@attribute Dimensions.Width numeric
@attribute 'Engine Information.Driveline'
@attribute 'Engine Information. Engine Type' {BMW 2.1L 4cylinders 200hp 208ft-lbs Turbo','BMW 3.6L 6cylinders 275hp 244ft-lbs','BMW 4.4L 8cylinders 350hp 326ft-lbs','BMW 3.7L 6cylinders 282hp 276ft-lbs','BMW 2.1L 4cylinders 212hp 260ft-lbs Turbo','HONDA 3.4L 6cylinders 282hp 255ft-lbs','HONDA 3.7L 6cylinders 307hp 275ft-lbs','Audi 4.3L 8cylinders 556hp 502ft-lbs Turbo', 'SUZUKI 3.0L 4cylinders 184hp 192ft-lbs','SUZUKI 5.4L 8cylinders 312hp 335ft-lbs FFV','SUZUKI 4.9L 8cylinders 282hp 296ft-lbs FFV','SUZUKI 6.1L 8cylinders 325hp 375ft-lbs FFV','SUZUKI 4.4L 6cylinders 196hp 262ft-lbs',
@data
140, 143, 202, 'all-wheel drive', 'BMW 3.3L 6cylinders 251hp 235ft-lbs', True, 6.6speed automatic select shift', 18, Gasoline, 25, 'automatic transmission', '2009 BMW A3 3.2', BMW, '2009 BMW Q3', 2010, 252, 235, 140, 143, 202, 'front wheel drive', 'BMW 2.0L 4cylinders 201hp 208ft-lbs Turbo', True, 6.6speed automatic Select shift', 22, Gasoline, 28, 'automatic transmission', '2009 BMW Q3 2.1T AT', BMW, '2010 BMW Q3', 2010, 201, 208

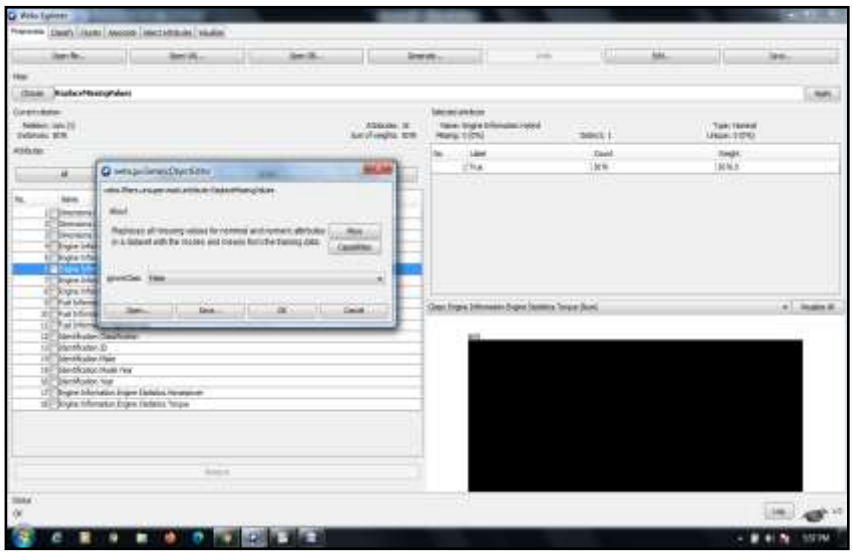

**Figure 5** Pre-Processing Cars Data

Figure 5 shows the pre-processing of cars data and by using replace missing value filter in WEKA can plugin to java platform. K-means clustering group is the similar features and values.

Let Z = {z1,z2,z3,……..,zn}be the set of data and w = {w1,w2,…….,vc} be the set of centers and choose 'c' cluster centers randomly. And calculate the distance between each data point and cluster centers and the centroid is important. K-means implementation is shown in Figure 6 and Figure 7 shows the results of farthest first clustering algorithms.

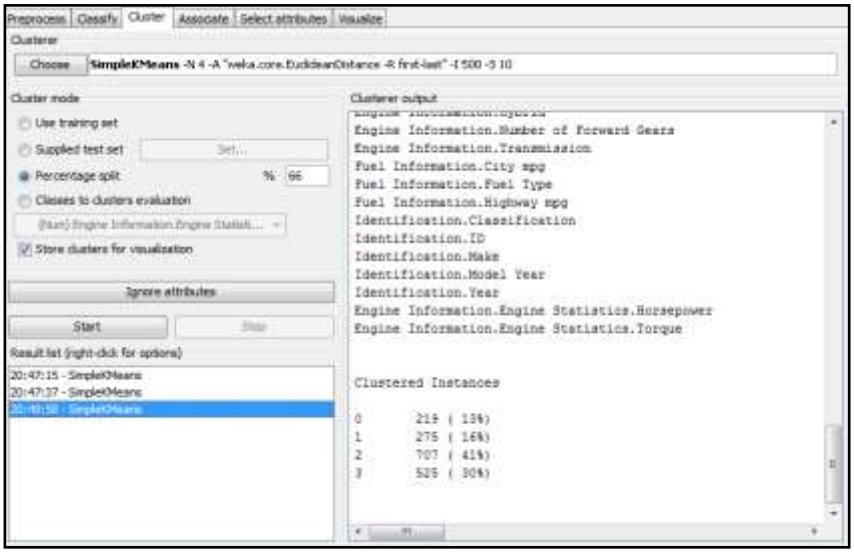


**Figure 6** K-Mean Clustering for Identifying Car Models

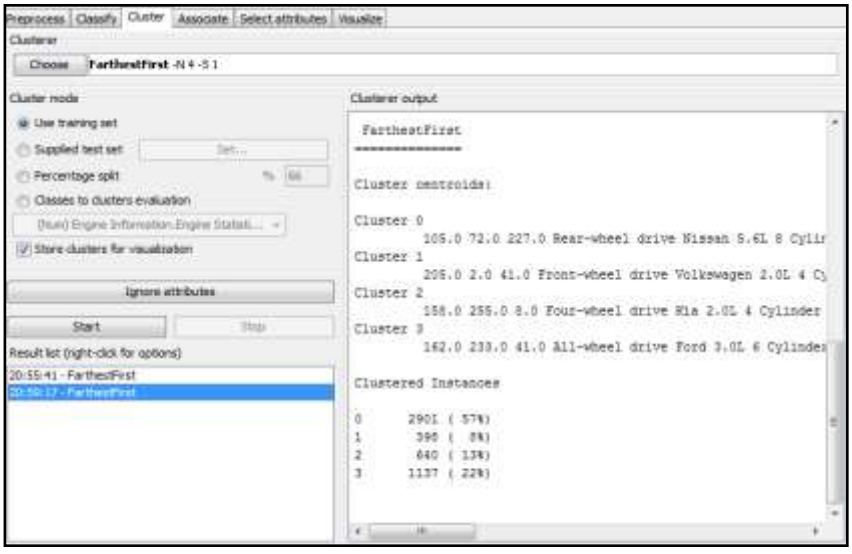


**Figure 7** Farthest First Clustering algorithm for Car with ARFF

k-Nearest Neighbor (kNN) make a prediction for an assessment attributes and in WEKA it is called Learner (IBK). It is utilized space measure to position k 'close' attributes in the training data for every test attributes and apply those chosen attributes to build a calculation. The benefit is that superior data of k offer soft to facilitate decreases the threat of over-fitting suitable to blast in the preparation data. The default measure is Euclidean distance and in WEKA can use Click as Add new tab in the algorithm side and then select the Choose button. Next get on IBK under the lazy group and click the OK button from the configuration of IBK and is represented by Figure 8.

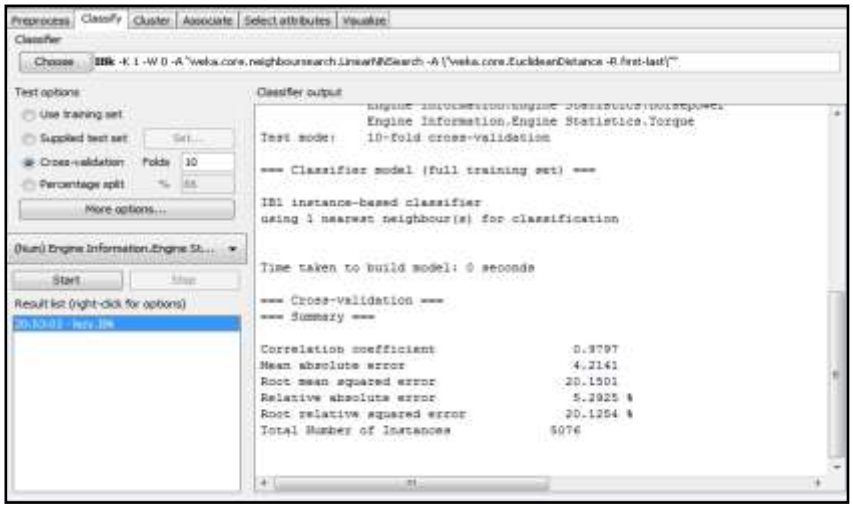


**Figure 8** Cars Data-Mining using IBk

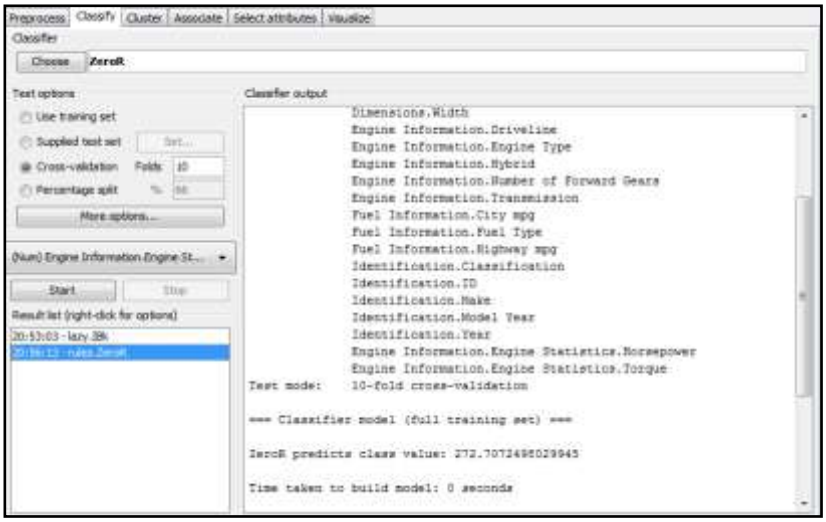


**Figure 9** Proposed System Classifications

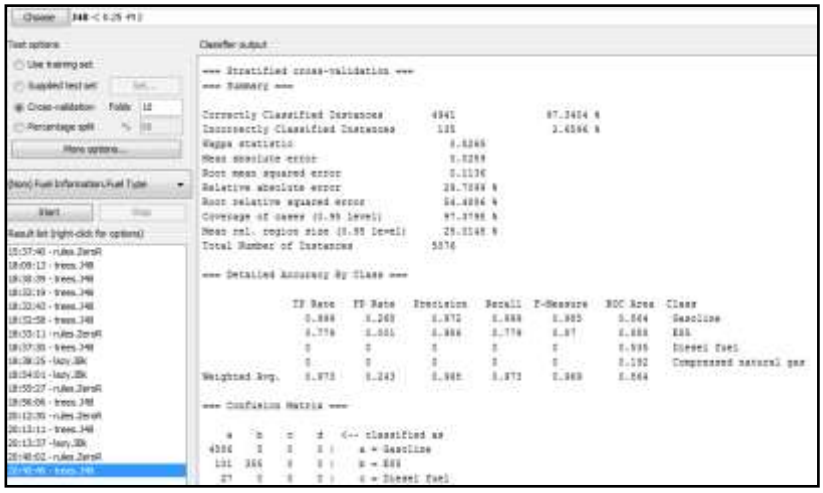

**Figure 10** Proposed Research Model

ZeroR is of the rule algorithm in machine learning. It is built in algorithm at WEKA as shown in Figure 9. The proposed research model with the classification results are shown in Figure 10.

J48-Decision tree algorithm:

Create_decision_tree
Input as length, width, driveline, engine type fuel type etc,
Data = load cars ()
Display data.target_name to predict the classes
Output: Tree structure for decision making
Model generation:
Create a node S:
If tuples in data partition are same class C, then
Return S as leaf node labeled with the class C
Apply instance selection method
Partition the tuples and grow subtrees for each partit
Extracting car data instances and target variable or class labels
Developing matrix format
Create training and testing data by the train_test_split or cros validation
Generate tree for each S
Return S:
Find the information gain by class label called fuel type.
Predict the accuracy of the classifier.

## 4 Results and Discussion

Data mining tool visualize the car behaviours with engine information, engine statistics, torque and model make as shown in the Figure 11 and Figure 12 respectively.

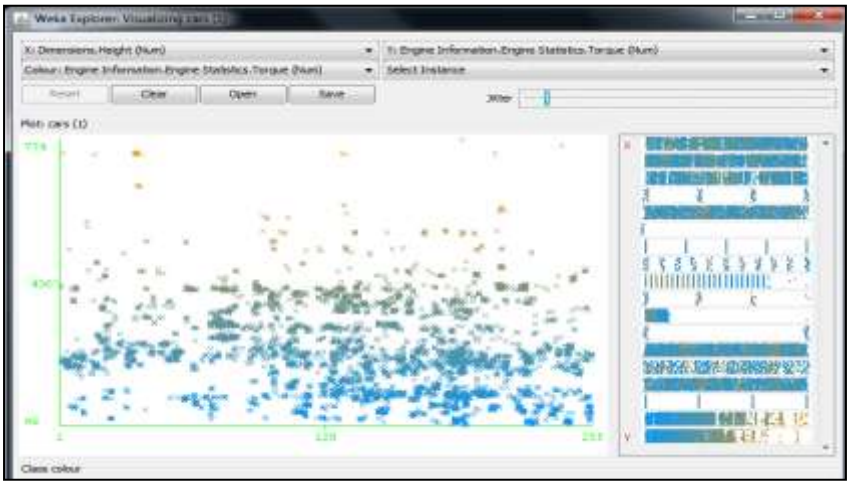


**Figure 11** Visualization of Result

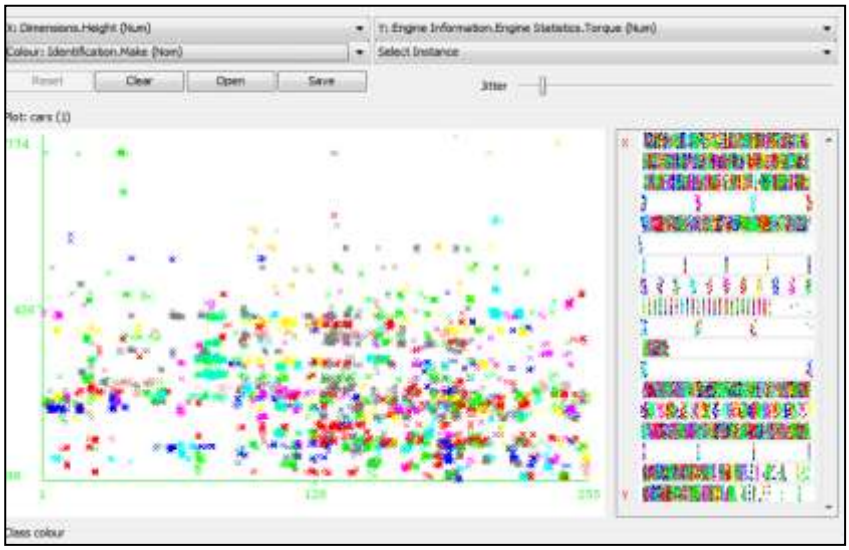


**Figure 12** Visualization of Result by Target Variables

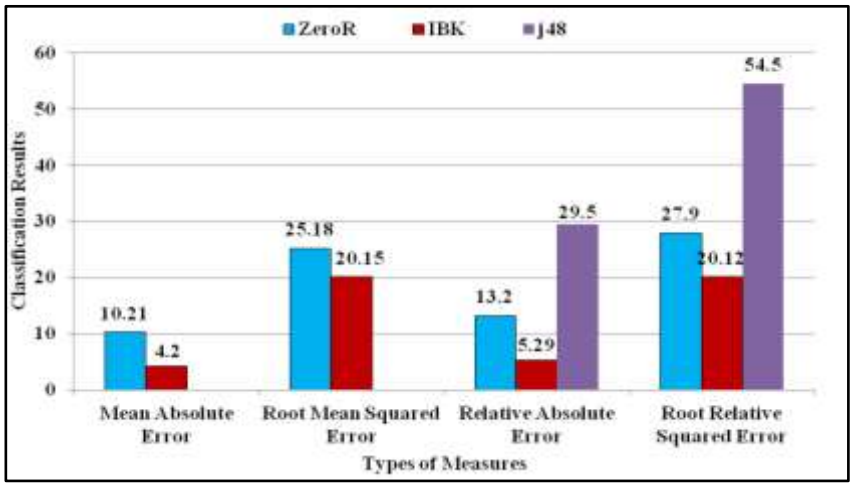


**Figure 13** Performance Comparisons of Proposed Classifier

Figure 13 shows various data mining performance measures for analysis of zeroR and KNN of IBK algorithm using WEKA. IBK produce good classification results compared to other algorithm in manufacturing car industries. Performance results of classifier when 6-fold cross validation used are given in Table 2. Among all the classifier J48 produces good result of 97.00 in class label as fuel information and fuel type. Similarly, performance results of classifier when 10-fold cross validation used are given in Table 3. Among all the classifiers J48 produces good result of 97.34 in class label as fuel information and fuel type.

**Table 2** Results of Classifier in 6-Fold Cross Validation

| Classifiers | Correctly Classified Instance (%) | Incorrectly Classified Instance (%) |
|---|---|---|
| ZeroR | 90.44 | 9.55 |
| IBK | 94.66 | 5.33 |
| J48 | 97 | 2.99 |

**Table 3** Results of Classifier in 10-Fold Cross Validation

| Classifiers | Correctly Classified Instance (%) | Incorrectly Classified Instance (%) |
|---|---|---|
| ZeroR | 90.44 | 9.55 |
| IBK | 95.15 | 4.86 |
| J48 | 97.34 | 2.65 |

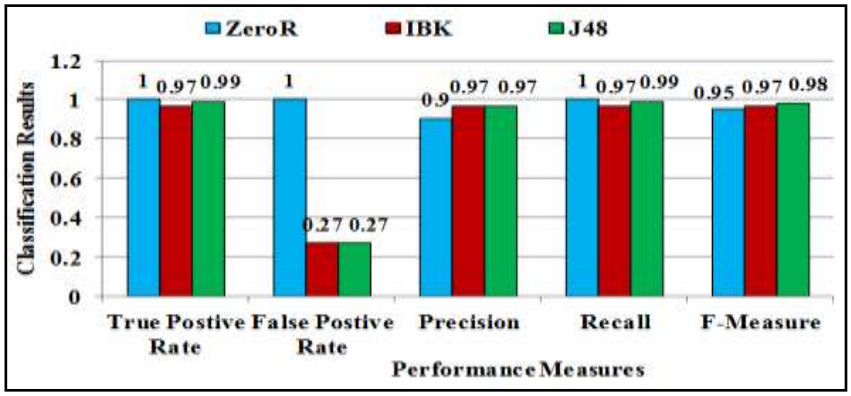


**Figure 14** Performance Comparative Results of ZeroR, IBK and J48

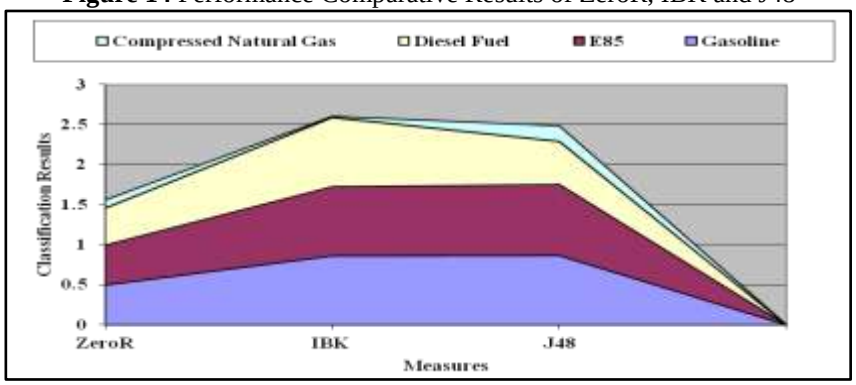


**Figure 15** ROC curve for proposed model

The performance results were compared with different classifiers of the precision, recall and F-measure as shown in Figure 14. The F-measure is calculated by 2 x precision + recall by precision x recall. By using this equation the value of precision could be calculated by number of correctly classified record by total number of records x 100. From the results it was observed that, J48 classifier improves the F-measure to the maximum level.
Figure 15 shows the results of Receiver Operating Characteristic (ROC) measures which represents sensitivity and specificity of the classifiers. The curves not near to closer shows good performance. From the ROC curve of our proposed model it was found that, fuel type gasoline, E85 and diesel fuel are produces good results to control fuel and the manufactures could focus on the customer point of interest.

## 5 Conclusions

This research illustrates the clustering algorithm of k-means and KNN in open source data mining tool with JAVA code. The training data cross

validation and various percentage splits are used to compare the classifier performance.

- The research assist the car purchasing pattern to the customer of view for manufacturer with help of clustering and fuel usage is improved by classification algorithm.
- Performance results of classifier while 6-fold cross validation J48 produces good result of 97.00 in class label as fuel information and fuel type.
- Similarly, performance results of classifier when 10-fold cross validation used J48 produces good result of 97.34 in class label as fuel information and fuel type.
- From the final results it was observed that, J48 classifier improves the F-measure to the maximum level.
- From the proposed ROC curve it was found that, fuel type gasoline, E85 and diesel fuel measures are produces good results with enhancement of sustainable development.

This research work can be extended to develop knowledge expert system for intelligent process analysis and performance monitoring in large scale manufacturing industries.

## References

[1] Carvajal Soto, J.A., Tavakolizadeh, F., Gyulai, D., "An Online Machine Learning Framework for Early Detection of Product Failures in an Industry 4.0 Context", International Journal of Computer Integrated Manufacturing, Vol. 32, pp. 452-465, 2019

[2] Bouferguene, A., Altaf, M.S., Liu, H., Al-Hussein, M., Yu, H., "Integrated Production Planning and Control System for a Panelized Home Prefabrication Facility using Simulation and RFID", Automation in Construction, Vol. 85, pp. 369-383, 2018.

[3] Samir Lamouri, Juan Pablo Usuga Cadavid, Bernard Grabot, Robert Pellerin, Arnaud Fortin, "Machine Learning Applied in Production Planning and Control: A State-of-the-Art in the Era of Industry 4.0", Journal of Intelligent Manufacturing, Vol. 31, pp. 1531-1558, 2020.

[4] Ji, W., Wang, L., "Big Data Analytics based Fault Prediction for Shop Floor Scheduling", Journal of Manufacturing Systems, Vol. 43, pp. 187-194, 2017.

[5] Dolgui, A. et al., "Data Mining-based Prediction of Manufacturing Situations Data Mining based", IFAC-PaperOnLine, Elsevier, Vol. 51, no. 11, pp. 316-321, 2018.

[6] Jurkovic, Z., Cukor, G., Brezocnik, M., Brajkovic, T., "A Comparison of Machine Learning Methods for Cutting Parameters Prediction in High Speed Turning Process", Journal of Intelligent Manufacturing, Vol. 29, No. 8, pp. 1683-1693, 2018.

[7] Y. Dong, S.J. Qin, "Dynamic Latent Variable Analytics for Process Operations and Control", Computers & Chemical Engineering, Vol. 114, pp. 69-80, 2018.
[8] C. Shang, F. Yang, B. Huang, D. Huang, "Recursive Slow Feature Analysis for Adaptive Monitoring of Industrial Processes", IEEE Trans. on Industrial Electronics, vol. 65, no. 11, pp. 8895-8905, 2018.
[9] Y. Dong, S.J. Qin, "A Novel Dynamic PCA Algorithm for Dynamic Data Modeling and Process Monitoring", Journal of Process Control, Vol. 67, pp. 1-11, 2018.
[10] Zhiqiangge, Zhihuan Song, Steven, X. Ding, Biao Huang, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning", IEEE Access, Vol. 5, pp. 20590-20616, 2017.
[11] Sheenam Jain, Vijay Kumar, "Garment Categorization using Data Mining Techniques", Symmetry, Vol. 184, no. 12, pp. 1-20, 2020.
[12] S. Nallusamy et al, "Sustainable Green Lean Manufacturing Practices in Small Scale Industries-A Case Study", Int. Journal of Applied Engg. 14 Soumitra et al. Research, vol. 10, no. 62, pp. 143-146, 2015.
[13] Fuster Parra, P., et al., "Ranking Features on Psychological Dynamics of Cooperative Team Work through Bayesian Networks", Symmetry, vol. 8, no. 34, pp. 12-17, 2016.
[14] M. Thangamani, V. Prasanna, "Data Analytics in Wine Datasets using WEKA Clustering", Int. Scientific Global Journal in Engineering, Science and Applied Research , vol. 1, no. 2, pp. 6-12, 2016.
[15] M. Thangamani, N. Suresh Kumar, "Effective Customer Patterns Analysis using Open Source WEKA Data Mining Tool", International Research Journal in Global Engineering and Sciences, vol. 1, no. 1, pp. 14-33, 2016.
[16] Danyang Sun, Fabien Leurent, Xiaoyan Xie, "Floating Car Data Mining: Identifying Vehicle Types on the Basis of Daily Usage Patterns", Trans. Research Procedia, vol. 47, pp. 147-154, 2020.
[17] K. Balakannan et al., "Performance Evaluation of Supply Chain and Logistics Management System using Balanced Score Card for Efficiency Enhancement in Indian Automotive Industries", Indian Journal of Science and Technology, vol. 9, no. 35, pp. 1-9, 2016.
[18] A. Kaleel Ahmed et al, "Study on Environmental Impact through Analysis of Big Data for Sustainable and Green Supply Chain Management", Int. Journal of Mechanical and Production Engineering Research and Development, vol. 8, no. 1, pp. 1245-1254, 2018.
[19] Hillol Kargupta, "Connected Cars: How Distributed Data Mining Is Changing the Next Generation of Vehicle Telematics Products", Lecture Notes, vol. 103, pp. 73-74, 2012.
[20] Kingsley Okoye, Syed Islam, Usman Naeem, "Ontology:Core Process Mining and Querying Enabling Tool", Ontology in Information Science, IntechOpen Publisher, 2018.

[21] Tin Kramberger, Bozidar Potocnik, "LSUN-Stanford Car Dataset: Enhancing Large Scale Car Image Datasets using Deep Learning for usage in GAN Training", Applied Science, Vol. 10, pp. 01-12, 2020

[22] Zhiqiang Ge et al., "Data Mining and Analytics in the Process Industry: The Role of Machine Learning", IEEE Access, Vol. 5, pp. 20590-20616, 2017.

## Biographies

**Soumitra Singh** is presently Chairman of Pixel Group of Technologies & Solutions dealing with IT, Education & Health sector. He is presently Research Scholar of FISLM, Jadavpur University also. He serves 19+ Years of different domains of higher academic, technical & health sector under different Universities like MAKAUT formerly WBUT, SMU etc. He completed his MBA in 2009 with live projects on MIS under SAIL, Govt. of India followed by ECE Engg, CICPI, CCHE. He delivered and registered 14 national and international conferences. He was awarded BHARAT VIDYA SHIROMANI by IIEM, New Delhi 2017. He was also served different committee members in International and National conferences.

**Partha Sarathi Chakraborty** is presently working as a Professor in the Department of Adult, Continuing Education & Extension, Jadavpur University. He has a teaching experience of around 16 years and Industrial experience of around 9 years. He has published more than 70 nos. research papers in National and International journals, apart from many more conference publications. He has successfully guided 4 nos. PhD candidates. He is also Editorial Board member of various journals.

**S. Nallusamy** did his B.E. (Mech.) in Madras University and completed M.E (Ind. Engg.) from Anna University, Chennai. He has received his Ph.D. from Jadavpur University, Kolkata in the year 2009. He has 27 years of teaching, research and administrative experience and has guided 11 Ph.D. and 36 PG scholars. He has published more than 180 research articles in reputed International Journals and has written six text books with ISBN. Currently he is working as Professor & Dean in the Department of Mechanical Engineering at Dr. M.G.R. Educational and Research Institute, Chennai 600095, Tamilnadu, India. His areas of interest are Composite Materials, Nano Materials, Material Science and Optimization.

**K. Balakannan** completed his B.E (Mechanical Engg.) in Madurai Kamaraj University in the year of May 1991 and M.E. (Engineering Design) in Bharathiar University. Received his Ph.D. from Jadavpur University, Kolkata in the year 2017. He has about 24 years of teaching and 3.5 years of industrial experience. He has published around 16 papers in international Journals and 5 papers in international Conferences. Currently he is working as Principal of Adhiparasakthi College of Engineering, Ranipet, Tamilnadu, India.