Indian Journal of Science and Technology, *Vol 8(15), IPL037, July 2015*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Manufacturing Quality Improvement with Data Mining Outlier Approach against Conventional Quality Measurements

R. Shankar[1*] and M. Sundararajan[2]

[1]Bharathiar University, Coimbatore - 641 046, Tamil Nadu, India; shankarramchanda@hotmail.com
[2]Department of Computer Science, Government Arts College, Nandanam, Chennai - 600035, India; drmsrajan23@yahoo.com

Abstract

Background/Objectives: This research study has been conducted in a scientific way to help manufacturing engineers and the management team to find out the hidden information from the data which are generated during the everyday manufacturing process. **Methods/Statistical analysis:** The methodology adopted in this activity is applying outlier analysis which is a data mining technique. The inter quartile range findings and analysis has been used here to find the hidden useful information from the process data with which a better insight could be established towards the improvement of quality of the product. The data used here have been collected from the automotive engine assembly and testing process. The study compares the results between the conventional and outlier analysis. **Findings:** The conventional style of checking and approving the engines based on the value pattern of Specific Fuel Consumption (SFC) which uses the design specification comparison with the actual data generally yields very minimal scope for the improvement of product quality in the perspective of design, Safety and reliability of the product because of the adherence of the same design specifications of the part drawings supplied by various suppliers. The competitive automotive manufacturing domain demands a different approach with which a better scope could be identified towards the improvement of product quality which is undoubtedly data mining. The outlier analysis using inter quartile range on the sample data of 500 engines revealed many important aspects where the improvement scope for quality has been identified as 15,000 Parts Per Million (PPM) against400 PPM of conventional quality analysis for the same data. **Improvement/Application:** This research is to offer an inclusive model hypothetically with actual data both for engineering and management people of manufacturing domain about the insights and benefits of employing data mining techniques towards the improvement of product quality with proven results.

Keywords: Automotive Engine Testing, Data Mining, Inter Quartile Range, Manufacturing Quality, Outlier Analysis

1. Introduction

The recent automotive manufacturing domain which involves high volumes of every day production has inevitably capable to produce huge volume of data during the manufacturing process. Generally the manufacturing process design involves uniform and even distribution of processes over the production line with sophisticated machineries where these processes are executed and the respective data are stored against the product identification numbers for the purpose of future traceability. The availability of such huge volume of data generated during the production process in a manufacturing organization of a product always tends the management of that organization to make use of the available process data to gain some business insights and useful results for the continuous improvements (KAIZENs) out of it in the perspective of return on investment due to the exorbitant

*Author for correspondence

investment done on the design and implementation of such huge data repositories and servers. The traditional approach with the available data in an organization is to analyze and study them towards getting an understanding on the results' parameters over a period of time or over a specified quantity of volume[1]. The general practice in the automotive manufacturing lines in India is to assign this activity of analysis and mining to the same engineer or supervisor who is responsible for the everyday production. This additional activity which is very cumbersome and also an add-on to their regular activities which drove them to find an easy way to perform this task i.e. Data Mining (DM). Now days all the global automotive manufacturers employ Data scientists and Data experts of the same domain to perform this activity towards finding better insights in business to render support to the top management people for decision making. DM technique towards the analysis of abundant data in an organization made an analyst to think in a varied way to proceed with an analysis on the irregular results or anomaly results towards achieving better results for which the analysts are employed for. In general, the minimum correction in one or some of the manufacturing process based on the anomaly or outlier would definitely result with an appreciable improvement over the majority stakes which is the driving force for this paper.

The testing process of an automotive engine is considered here for study. In general there are minimum 15 parameters are monitored in the process of testing an automotive engine as part of its manufacturing process. Testing is the important process which decides the engines' movement to the immediate customer. In this paper a set of 500 engine test result data of one parameter has been taken for outlier analysis and the same has been analyzed to find GOOD and BAD outliers in relation with the conventional analysis and also to identify the best and worst ones to use for further analysis towards the improvement in product quality and to analyze the aspects/scope for the quality improvement of the product using a typical automotive Engines' Test result data with the adoption of data mining outlier analysis and relate the same with conventional measurements and to bring out the benefits of this kind of approach in Manufacturing.

2. Materials and Methods

The use of DM in improving the manufacturing Quality has widely been discussed now days due to the advanced technological improvements in the field of manufacturing. Referencing[2] authors emphasize the importance and the necessity of DM techniques towards the identification of the hidden patterns / abundant information from the data which are produced during manufacturing activity. They specifically focus on mining the quality related data by the use of classifier models towards the achievement of improvement in manufacturing quality.

Same way the application of DM Outlier analysis is employed elsewhere[3] where univariate datasets were taken and analyzed using the methods: 2 SD Method, 3 SD Method, TUKEY's Method (1.5 IQR), TUKEY's Method (3 IQR), Adjusted Boxplot, 2 MADe Method, 3 MADe Method and Median Rule. The results obtained were compared using all the methods with the same univariate datasets. It reveals that all of them show almost similar patterns in an overall view. Also the total average percentage of outliers for the adjusted boxplot using TUKEY's method seldom changes, even decreases very slightly, when data are skewed. In practical the use of multiple methods in a comprehensive way would be suitable for outlier analysis[4].

Further work[10] carriedout with the same Outlier analysis on the real-time data from ten different datasets using IQR method to detect outlier automatically as a data preprocessing activity. The work mainly focused on the identification is mainly for the automatic elimination purpose. The outliers can be assigned specifically based on the requirement for which the analysis is meant for. Though the methodology, models and the basic data types for the outlier calculation is of manifolds, the correct arrival of correlation of the outliers mainly depends on the domains requirement[5].

Considering all these previous reports in this paper where the sample data is almost tenfold. The data considered here is the real-time gathered data from the successfully running engine model of an automotive manufacturing company where they are in real search of an innovative means to find out the increased scope for improvement from the manufacturing process data results analysis.

3. Automotive Engine Manufacturing Process

The Figure 1 is the simple representation of an automotive

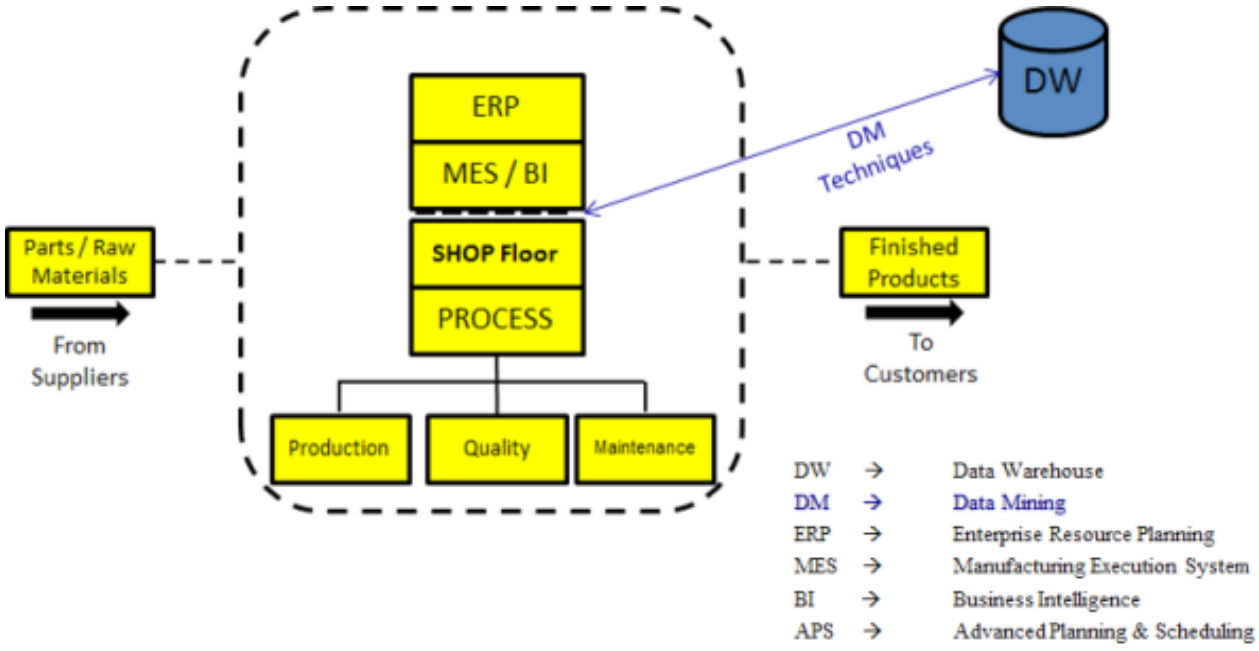

Figure 1. Automotive Engine Manufacturing Process.

engine manufacturing process where the execution of the assembly process of the engine in the shop floor which is integrated with Enterprise Resource Planning (ERP) and Manufacturing Execution Systems (MES) and the process gets completed first based on its design specifications by using the parts from various suppliers received as per the design. The volume of data that are generated during the production process is moved and stored in the data repositories. The TESTING of an engine would always be the final process where the engine is made to Crank ON - checked and validated for various parameters like Power, Torque, and Specific Fuel Consumption (SFC) etc., for which the engine has been designed for. Such validated results are stored in repositories against the allotted engines' serial number for easy genealogy and traceability.

4. Data Mining-outlier Analysis

The process or practice of examining large pre-existing databases in order to find out new/hidden information is called Data Mining (DM). The need for DM tools and techniques are increasing every day in manufacturing sectors because the latest advancements in the machine building technology enables most of the manufacturing process equipment and machineries compatible to Data Mining due to the use of sophisticated process control systems. Sensors keep track of and store process parameters such as pressures, flow rates temperatures, efficiencies and other variables that are needed to monitor the process. Thus huge amounts of data are collected and stored[1].

Hawkins states that an outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism[2]. Outliers can arise from several different mechanisms or causes. ANSCOMBE (1960) sorts outliers into two major categories: those arising from errors in the data, and those arising from the inherent variability of the data[3]. Not all outliers are illegitimate contaminants, and not all illegitimate scores show up as outliers[4]. It is therefore important to consider the range of causes that may be responsible for outliers in a given data set[8]. What should be done about an outlying data point is at least partly a function of the inferred cause.

5. Discussion

The detailed illustrations done on sample data of 500 tested engines have been described in this session. The corroboration done using the data samples in the con-

ventional style and the application of DM outlier analysis on the same sample data is described here. Primarily the results have been analyzed conventionally and same data are examined with outlier analysis as explained in sessions 5.1 and 5.2 respectively. The main thing which differentiates both the evaluation is the factor of Mean and Quartile range. These kind of evaluation based on the DM techniques definitely reveals the really unexpected/hidden information which in other term can also be named as intelligence inside the data. This is the key results of outputs which enable the management to approach design parameters and arrival of design specifications with a novel and innovative style. Making the quality specifications of a product to a high standard which is a mandatory need of current competitive automotive manufacturing environment is highly possible only with these aspects of applying DM techniques on the data where a new dimension could be arrived in fixing the design specifications. The very strict design specification is also of no use when it deviates from the concept of "Design for Manufacturing (DFM)" i.e. the specifications which would be not possible to meet in practical with the available manufacturing technologies. DFM simply reiterates that the cost and complexity involved for the given product to be manufactured within the given specification scales to befit as minimum as possible which needs to be taken care by the designers. Here this discussion session illustrates the GOOD and BAD outliers because of the above said consideration of DFM.

6. Conventional Analysis on the Sample Data

A sample of 500 automotive engine testing results from an automotive engine manufacturing organization has been taken for study. The main objective of this analysis is to grab the attention of people who involved in the manufacturing process to find out the root cause for the anomalies in engines' Specific Fuel Consumption (SFC) Data found during testing process of the engines during their manufacturing cycle. The process of identifying outliers with sample volume of data resulted as follows during the conventional quality measurement process.

The Scatter graph plotted with SFC test results data conventionally reveals that only 2/500 data which is almost 0.4 % (above green line in Figure 2) is out of specification i.e. in this SFC case it can be said that the above are the data that are out of specification which in turn means for the manufacturer to take it for rectification process to make those products to bring within the specified SFC limits which is a cumbersome and cost involving factor. The Figure 3 Above reveals the result of the conventional analysis on the data to monitor the SFC quality level.

Specification:→ SFC < 233 g / kWh

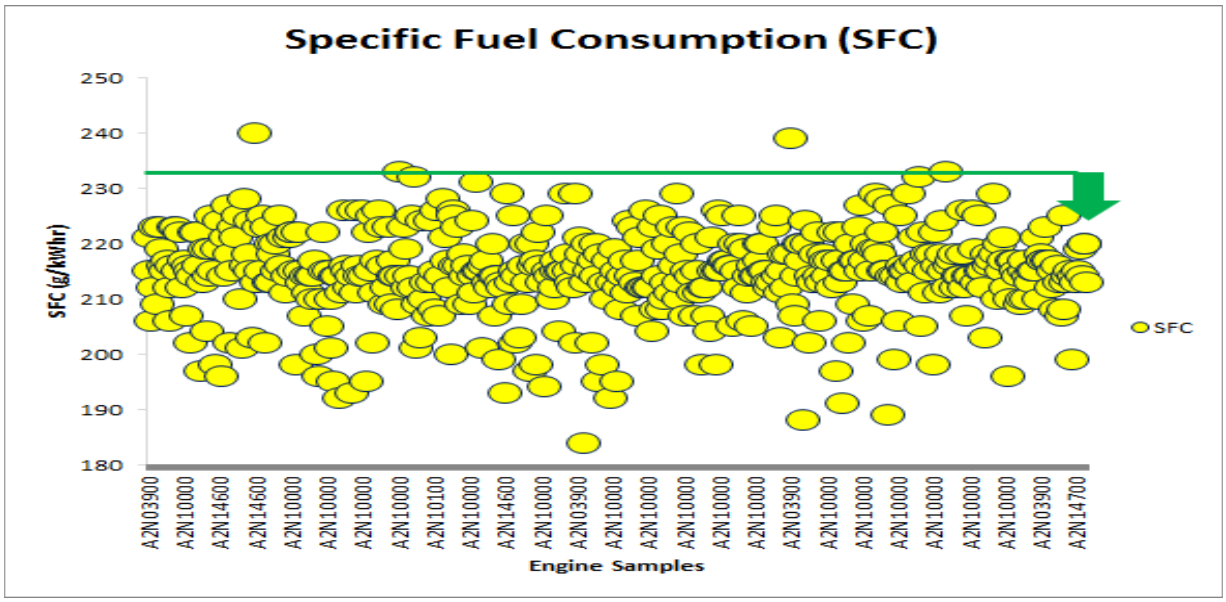

Figure 2. Conventional style of checking the SFC data of Engines for Quality analysis.

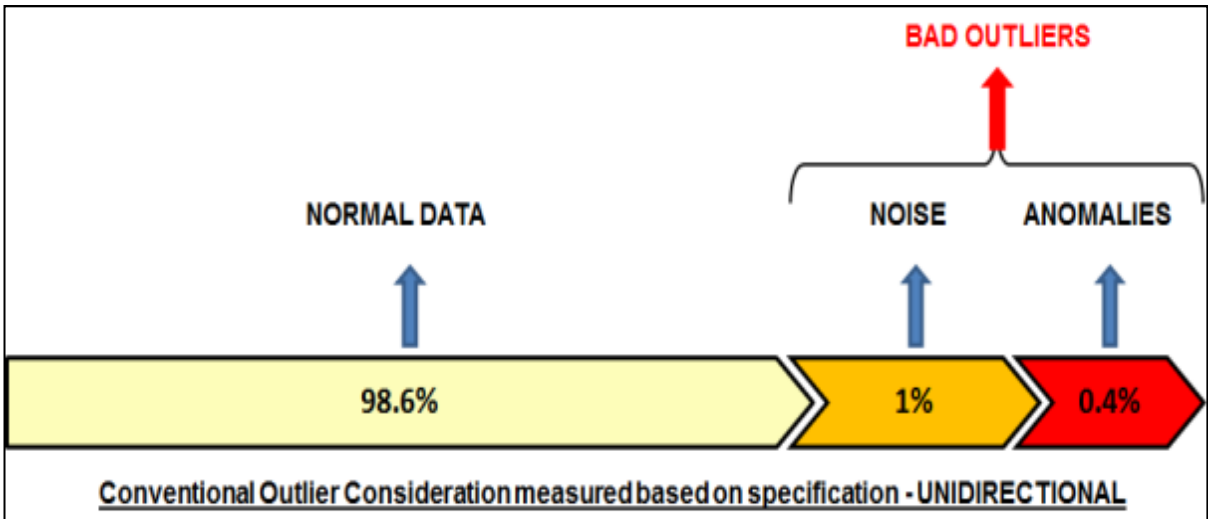

Figure 3. Result of conventional measurement of SFC Quality data

The regular course of action in a manufacturing firm might give seldom attention towards the quality improvement by using anomaly result which is only 0.4% (2 of 500 engines) because the intention of the manufacturing engineer is to give the product within the specified SFC limits due to the exorbitant timelines and product delivery pressure.

The recent automotive engine testing equipment systems are intelligent enough to act as a quality gate and to reject the tested engine which fails to meet the design specifications with special prompts to the engineers and alerts the manufacturing engineers to keep them aside for rectification activities through which the defective product is prevented to reach the customer. Hence the process data of such defective engines are also made to store in the system and the same engine when undergoes retest after rectification of the highlighted problem where the same engine's retest data are also stored separately for the genealogy and traceability purposes. But, for grabbing the attention of the management towards the improvement to be done on,

- Quality of the product by zeroing/analysing anomalies
- Reducing the Inventory cost of such anomaly engines that would lie aside for salvaging activities.

Salvages are additional activities of lesser interest in manufacturing plants which shall be turned up as an interesting activity by the use of a strong data-mining analysis only with which the possibility lies to bring out the hidden interesting factors and patterns from the available data. On the same set of sample data, the analysis performed using Outlier analysis of DM technique is described below in chapter 5.2.

6.1 Outlier Calculation using Inter Quartile Range

In any set of data it is important to be on the lookout for outliers. Outliers are individual values that fall outside of the overall pattern of the rest of the data. This definition is somewhat vague and subjective, so it is helpful to have a rule to help in considering if a data point truly is an outlier. One rule that is very simple to apply is Inter Quartile Rule (IQR) which utilizes the interquartile range[6]. The IQR follows certain steps in calculating outliers which are stated below,

i. Arrange data in order.
ii. Calculate first quartile (Q1)
iii. Calculate third quartile (Q3)
iv. Calculate inter quartile range (IQR) = Q3-Q1
*v. Calculate lower boundary = Q1-(1.5*IQR)*
*vi. Calculate upper boundary = Q3+(1.5*IQR)*
vii. Anything outside the lower and upper boundary is an outlier.

The same sample data which was used to plot the graph in Figure 2. has been arranged in ascending order to find the median, first quartile, third quartile, inter quartile range and the boundaries, the necessary details as per the IQR steps have been performed on the sample data and the results obtained are as follows,

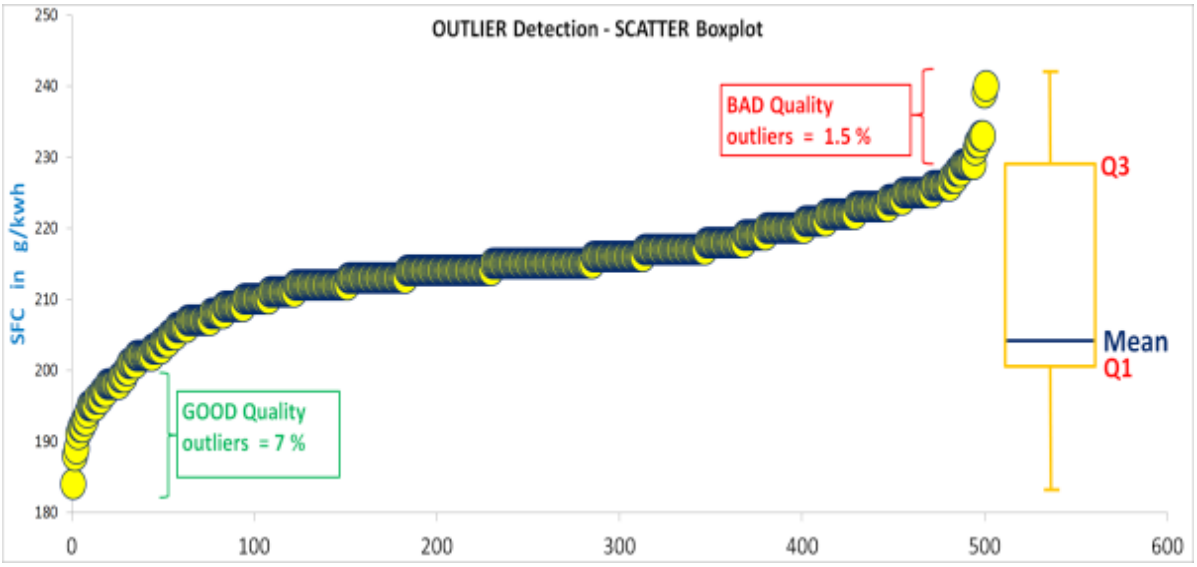

Figure 4. Result of OUTLIER analysis and BOXPLOT of SFC data.

Median Q2 = 215
First Quartile Q1 = 212
Third Quartile Q3 = 219
Inter Quartile Range IQR = 1.5*(Q3 – Q1) = 10.5
Tukey's Fence F = Q1 – 10.5 ~ Q3 + 10.5 = 210.5 ~ 229.5

The Figure 4 shows scatter boxplot diagram based on the obtained result of the Outlier analysis performed on the same samples of SFC data of the engines. The Q1, Q3 and Mean values are indicated with Orange and blue lines respectively. The boxplot of the scatter reveals many interesting factors like the relativity of mean data with the given SFC design data specification, flow of outliers in bidirectional way, classification of good and bad outliers etc., from the view point of an engine manufacturing engineer. Surprisingly the outcome of the graph revealed many interesting hidden details which are explained below. The Figure 5. Represents the result in visually understandable form in which the good and bad outliers shown in bidirectional spread in green and red colors which are 7% and 1.5% respectively.

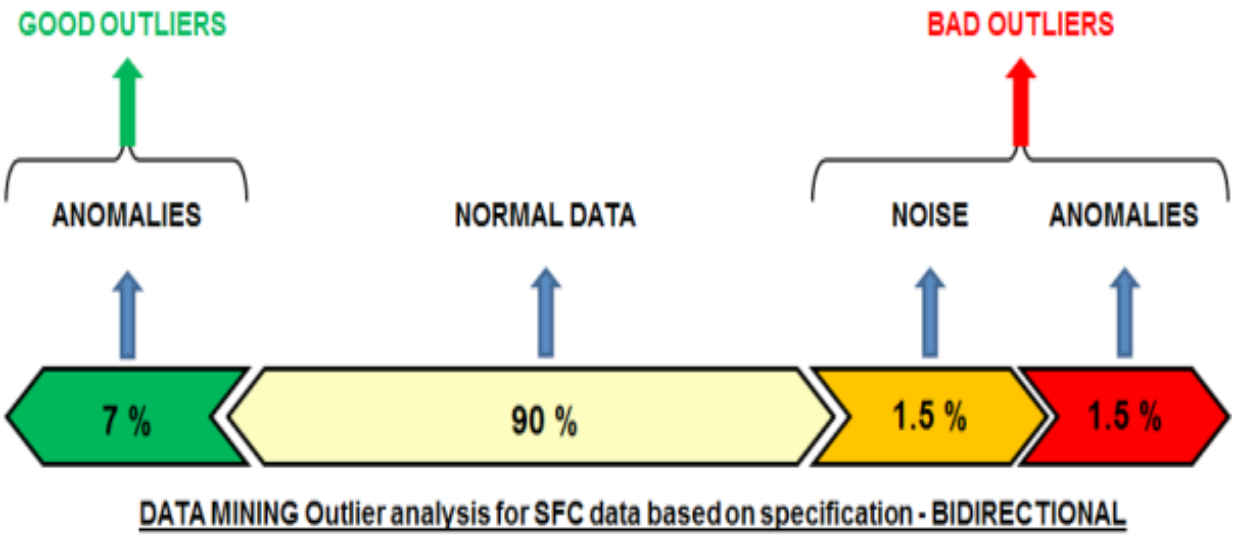

Figure 5. Result of OUTLIER analysis of SFC Quality data.

The BIDIRECTIONAL findings of Outliers identified from the data sample which helps in benchmarking the activities to be planned towards the improvement on Quality .i.e. the things which needs to be done are planned from the samples of GOOD outliers (7%) and the things which are to be avoided are planned from the BAD outliers. This is an important outcome from the outlier analysis which cannot be thought of from the conventional analysis. (Here the associated parameters of SFC during engine testing such as Temperature of coolant, ambient temperature, and Torque and Power values of the same engine at the same specified speed are adjudged to be the factors for consideration while deciding for the plan of activities that are to be done as a consequence of this result)

The identification of mean of data sample which is 212g/kwh which is fairly lower than the SFC specification under which 50% of the data lies with. This is another hidden result with which the Engine development team gets ample space/scope of work in improving fuel economy with a minimum target of 10% which is very high industry potential in automotive business. (This aspect of availability of 10% scope[12] towards the improvement of fuel economy needs to be understood with keeping other factors constant and based on the fuel consumption specification which is 233g/kwh against the mean value 212g/kwh i.e. there is scope to get the same fuel economy with the lesser consumption of fuel than the specified limit which would definitely act as Unique Selling Point - USP in the automotive business especially in commercial vehicle sales)

With respect to the given SFC specification the BAD outliers are about 1.5% in which half of them can still be declared as PASS products. As a whole the Outlier analysis reveals that there are scope in both GOOD and BAD directions (BIDIRECTIONAL) of the given design specification for the improvement in SFC quality.

7. Conclusion

The severity of the anomaly found using conventional methodologies is 0.4% which in turn reveals that the first pass ratio is 99.6 % which is a major proud factor of the manufacturing perspective. The same gets contradictory in terms of scope for quality improvement when it is analyzed using data mining technique of Outlier analysis which revealed 1.5 % outliers which is an indication to the top management to get a clear business insight about the SFC quality. The DM approach on the data also gives an idea to the top management that there exists a superior analytic approach towards quality improvement on the product. QCT (QUALITY, COST and TIME) improvement in any form is being given top priority in the manufacturing organizations due to the tough market competitions and increasing manufacturing overheads. Today's facilities in many manufacturing organizations have all the capabilities inbuilt which are required to get connected with Data – mining systems. The requirement in the organization is the non-availability of such Data – mining systems / practicesand the lack of data analyst who should possess the domain knowledge to customize the data – mining systems to provide such business insights day – to – day and time - to – time to the decision making authorities. The above analysis is further taken in relation with association rule mining and clustering analysis with the same set of data to further reveal the hidden information with respect to the quality aspects of the product.

8. References

1. Raggatt PR. Duplicates or singletons?—an analysis of the need for replication in immunoassay and a computer program to calculate the distribution of outliers, error rate and the precision profile from assay duplicates. Annals of Clinical Biochemistry: An international journal of biochemistry in medicine. 1989; 26(1):26–37.
2. Hawkins D. Identification of outliers. Chapman and Hall London; 1980.
3. Anscombe, FJ. Rejection of Outliers. Techno Metrics. 1960; 2:123–47.
4. Barnett V, Lewis T. Outliers in statistical data. 3rd edition. New York: Wiley; 1994.
5. Chen ZY. Study on Sample Outliers Statistical Test Methods. 2015 International Conference on Social Science and Technology Education; Atlantis Press; 2015.
6. Hall M, et al. The WEKA data mining software: an update. ACM SIGKDD explorations newsletter. 2009; 11(1)10–8.
7. Carroll CT. Six Sigma for powerful improvement: a green belt dmaic training system with software tools and a 25-lesson Course. CRC Press; 2013.
8. Han J, Kamber M, Pei J. Data Mining Concepts and Technique. 3rd Edition. USA: Morgan Kaufmann publishing; p. 443–5.
9. Seo S. A review and comparison of methods for detecting outliers in univarate datasets. University of Pittsburgh; 2006. p. 30–5.

10. Sunitha L, Balraju M, Sasikiran J, Venkat Ramana E. Data mining: manufacturing and service applications. International Journal of Advanced Research in Computer and Communication Engineering. 2014; 3(6):7255–7, ISSN(online): 2278-1021, ISSN (Print): 2319-5940.
11. Rokach L, Maimon O. Data mining for improving the quality of manufacturing: a featureset decomposition approach. Journal of Intelligent Manufacturing. 2006; 17(3):285–99.
12. Raggul N, Saraswathi R. Design and development of secondary clarifier for paper and pulp industry with a case study. Indian Journal of Science and Technology. 2014; 7(12):1939–49
13. Aggarwal CC. Outlier analysis. Springer Science & Business Media; 2013.