MERIT RESEARCH JOURNALS
www.meritresearchjournals.org

Merit Research Journal of Engineering, Pure and Applied Sciences Vol. 5(2) pp. 019-024, October, 2019
Available online http://www.meritresearchjournals.org/epas/index.htm

Original Research Article

The application of data analysis algorithm in quality control, case study research: the body making hall of Peugeot 206

Mohammad Ali Afshar Kazemi[1], Mohammad Mehdizadeh[2], Ali Reza Agha Gholizadeh Sayar[3] and Azam Ali Khademi*[4]

[1]PH.D, Central Tehran Branch, Islamic Azad University, Tehran, Iran

[2]PH.D,Science and Research Tehran Branch, Islamic Azad University, Tehran, Iran

[3]PH.D,Science and Research Tehran Branch, Islamic Azad University, Tehran, Iran

[4]MSC, Central Tehran Branch, Islamic Azad University, Tehran, Iran

***Corresponding Author's E-mail: a_alikhademi@yahoo.com**

Abstract

This study investigated the application of data analysis algorithm in quality control, case study research: the body making hall of Peugeot 206. In recent decades, human ability for production and collection of data has been increasing rapidly. Some factors play important role in this process; for instance, widespread usage of computer abilities in diverse sciences, development in instruments for collection data, scan and coherent information systems, coherent banking systems, and electronic trading. The explosion development in saved data caused to produce new technologies and automatic instruments in order to convert the numerous content of data to science and information. Therefore, data analysis is a solution for the problems. Data analysis is predicated to use of data analysis instruments in order to detect the patterns and valid relations which has unknown so far. The present study is applied to report data analysis in a production hall of Iran Automotive company; which the main aim is to indicate the valid assessment and amount of reliability and accuracy of inspectors in the body making hall. In the present study, the researcher attempts to present this accuracy by percent and using CLEMENTINE program, and attempts to illustrate it when an inspector confirms a body, to what extent could have the probability to require again inspection by the editors immediate reaction system. In the present study, it has utilized the primitive data from information systems about the quality which were access in order to accomplish computation by CLEMENTINE program. The findings of the research revealed that the prediction is trustable in 85% of occasions; by using data analysis in the process of quality control and anticipation the accuracy of people operation in the body making hall and preventing of exploration of the problem and again defect in reactivity section.

Keywords: Data analysis, process of quality control

INTRODUCTION

At past, exploitation of useful information from saved data was usually done handy, and by the analysts. As regards that handy analysis was slow, expensive, and also the content and complexity of data were increased every day, the handy analysis moved to the indirect analysis and using the computer methods (Azar, 2011). At recent two decades, human technical abilities have rapidly increased in production and storage data. The amount and extent of data sources are increased rapidly; and some of these data sources have developed in the way that include several hundred million or several milliard have re-corded. However, data content in sources has increased

momentarily according to the coherent information systems, coherent banking systems, and electronic trading; and caused to create numerous data sources. The necessity to exploit the science and knowledge from data sources has been increased; and data analysis is introduced as one solution for the problems. Data analysis is figured on as one of the most powerful instruments for analysis, because of access a numerous content of data and hidden science and knowledge inside of it. The achieved science and knowledge is applied spaciously in trading management, control of production and analysis of market, engineering design, and scientific researches. Therefore, it is necessary to utilize the nascent technology and automatic instruments in intelligence way in order to convert the numerous data to science and knowledge. So it is necessary to utilize of information technology for applying this science; and data analysis is proper solution to exploit these properties (Mahmoodi et al., 20014). According to data sources in generative factories, the aim of the present research is to utilize the inscribed and topical information of quality control include, quality control of produced cars from body making hall, the determination of defect of cars which were probed, the number of the body, the time and date of the control, the person who controlled with exclusive code, the section which is probed, and the others; in order to increase the quality factor, and all of this process is carried out by utilization of data analysis. Hence, the other aims of the present study are to assess of validity and indicate the amount of reliability and accuracy of people (COM) and production inspectors (ACC). Moreover, In the present research scrutinizes the background of data analysis in generative organizations and companies, in direction to create the requirement knowledge, the researchers have utilized an applied sample which were about a data analysis project in Iran Automotive in order to solve the problem of a numerous content of saved daily information and not utilization of them. As a result, the findings of the research were used in order to increase quality control of generated production and solve the problem of dissatisfaction of costumers. The history of detecting the knowledge from information sources does not have a very long antiquity; and today it is known as data analysis. The term of detecting science was introduced for the first time in 1990s; and at last the researchers' concentration was interested in the data analysis algorithm. The aim of data analysis is to detect nascent, reliable, and pursuable science, which can be pursued by the use of instruments of artificial intelligence and statistics in numerous content of data (Hand et al., 2000). Data analysis is predicated the interesting of data analysis instruments in order to detect valid relations and patterns which are unknown up to now. These instruments can be statistical patterns, numerical algorithm, and learning methods. Data analysis is not used for collection and management of data exclusively, rather than it consists of analysis and anticipation of information. Data analysis or detection of knowledge in data sources is the exploitation of useful and practical information from unknown data. This exploitation consists of some technical method for instance, categorization, summarization, learning rules, finding the relationship among networks, analysis of variations, and detection of irregularities (Fayyad et al., 1996). In other words, data analysis is to detect noticeable, unexpected, and valuable structures from spacious collection of data; and it corresponds to statistics and data analysis completely (Hand et al., 2000). The following is illustrated the procedure of detection of data sources (Despand, Taker, 2010). Choice of data: Data related to analysis and making decision and separated from the other data. Preprocessing of information: In this process, data is processed and made coherent. Conversion of data: The chosen data is converted to proper form for data analysis process. Data analysis: In this process is used an intelligence method to exploit applied and useful patterns and then make decision on these methods. Assessment and interpretation: In this process, noticeable patterns are known and presented to the users. The utilization of visualization methods in order to facilitate the users is necessary. An anticipated data analysis presents a model of system which consists of distinctive data. The main aim of it is to create a pattern for categorization, anticipation and estimation of data. Descriptive data analysis present the new and non-trivial according to collection of available data; and the main aim of it is to anticipate analyzed systems by the use of available ties and patterns. It is important to assign that data analysis comprises of several fields. The technologies that data analysis interests of them include: statistics, machine learning, optimization methods, methods of distinction of pattern, information bank, imagination making, neural network, mathematical models, retrieval of information, genes algorithm, artificial intelligence (Chapman et al., 2000). Data analysis is a process of detecting of ties, patterns, and meaningful and new process; that to probe numerous content of saved information in data sources by using the technology of distinction of pattern (Han and Kamber, 2001). The instruments of data analysis analyzed the data; and detected the pattern of data that can be used to determine the strategy of trading, knowledge source, scientific and medicine researches. As it was mentioned previously, data analysis has great potential that can be used in researches and studies. The idea of utilization of collected data in diverse direction of primitive aims is developing in order to anticipate and exploit the available and hidden knowledge of data. But limited researches and studied have carried out in data analysis applied in the process of construction, production, and quality up to now; that the majority of these researches are related to special applying of these techniques in electronic sources and semiconductor (Harding et al., 2006).

As regards that industrial organizations and companies confronted with numerous data and information about the quality of production, dimension and tolerance of smithereens, demanding of the market, staff, and material sources; and these data can include valuable information and patterns. Therefore, it seems that one of the most important applications of data analysis is in industrial companies. Today, there is spacious information banking from production and securement up to marketing and sale in these companies that include the information about the features of primitive material, generative productions and the others. Findings hidden knowledge and science of information can assist the industrial determiners. The utilization of advanced data analysis techniques in macro level can be applied; for instance, categorization, cluster; that can apply for categorization of industrial companies, finding of special and valuable pattern in successful companies, finding a pattern in order to achieve strategic goals, finding the critical points in financial management of companies, and the others (Neaga and Harding, 2005). Accomplished researches in recent years in global level, introduced the possibility of the applied of data analysis techniques in generative organization in fields of conservation and reparation (Romanowski and Nagi, 2001), identification of fault, improvement of quality (Last and Kandel, 2001) and the others. It is obvious that these applications can be developed in other processes of the production in organization like schematization process and the control of production quality. In following of the study is presented terse explanations about the company of Iran Automotive, schematization process and the control of production quality in this company for more acquaintance to the topic of the research. According to the license of economic administer (industries and minerals) date 25th of March in 1342 was permitted to Iran National (Iran Automotive) to launch in manufacturing Peykan car. In 1365 the permission of foundation of manufacturing of diverse car riding company in all types of 4 cylinders was given to the company in date 1365. The installation of Peykan car making was established in production section by primary production of six thousand per year in 1346. After establishment, the first cars riding produced by the credit of in 1346. The first car which was presented to the market in this year was Peykan in two models, Deluxe and Carluxe. The industrial company of Iran Automotive is the greatest company for car production in Iran and Middle East; that it has been dedicated to itself the average of 50 to 55% of Iran production cars permanently (Iran Automotive company official Site). The system of assessment and verification of production should be capable to assess the changes and modulation of quality level of production and also inform quickly in order to make proper decision, as a representative of costumers. In order to determine and assess the level of production, it is utilized the systems of quality aiming. The master manager of the company utilizes the systems of management information (MIS) of company in order to inform the condition of quality of diverse production constantly; and has the authority to carry out the proper dissuasive operatives if there are any sensible differences. The present research is about the operation of the final inspector in production and also the quality in body making section of Peugeot 206 (Neaga and Harding, 2005). In the body section 5 (which is producing now in two shifts), the production inspector (ACC) functions to control totally produced body after completion of all process at the end of the Calvabarline (the last line of body production in body making section of 206). Then if there is a fault in it, the inspector has to inscribe it in body card and also inform the respective responsible person, in order to eliminate the fault, and at the end the body was approved. Finally the body without any fault will transfer to the final station or COM station which the inspectors of quality resided there. The inspectors adapt the quality of the body with the expectation of costumers, and finally the body is accepted or not accepted. If the body is not accepted, it would be informed indirectly, and it should be reworked again, which is treated as negative point for production. Each person of COM and ACC has the seal with special number which belongs to him; for instance the number of seal 002 is belonging to Mr. X.

The production sections according to their process in Iran Automotive Company are as follow:

Smash section
Body making section
Color section
Montage section
Supplementation section

In production sections of Iran Automotive, body hall is the costumer of smash section, color section is the costumer of body, the montage section is the costumer of color, and the supplementation is the costumer of montage section. In order to prevent quickly of any fault transmission to the costumer section, a system is utilized as immediately reaction. In this system, when a section was observed a fault of previous section would inform it to the immediate reaction of that section. The representative of the section refers to the costumer section in order to inform the fault, and to observe it, and to distinct the reason, and finally in order to transfer the fault information to the source section. If the representative of the source section approved it, the analyzed sheet is presented to the representative. Therefore, the fault is transferred to the source section, but if it is not accepted, it is referred to the central analyzed unite, and in this section will announce the final result as a commandment.

METHODOLOGY

The present study, according to the aim is considered an

applied _ research, and according to collection of data is considered attributive _research. In order to achieve the information of section of quality control, it is utilized the inscribed information from data sources. The type of data in this research is in the case of numerical (continuous), and nominal (discrete). In the present study, it is utilized the standard model CRISP_DM; which was a standard method for data analysis, and it was established in late 1996 by three huge companies, DAimelerKelimer, S.P.S.S.One.C.R.N (Chapman et al., 2000).

This method introduced a process model for data analysis; which is a review of circulation of life of every data analysis project. The circulation of life of a data analysis project of this method consists of six stages: the apperception of trading problem, the apperception of data, model, assessment of results, and application of the model. The utilization of CRISP_DM, As regards that it is a universal standard for administration of research process; below the applied structure of the research according to the schedule of the standard is illustrated:

Apperception of trading problem

The present stage is focused on the aims and requirements of the project from trading point of view that the most aim of trading of car making is to present qualitative production. Then the conversion of this knowledge to a definition of data analysis and designed schedule to achieve the aims. The aim of data analysis in the present study is classification which is predictor type.

Apperception of data

The present stage is launched by collection of data; and then occupied to description and determination of the data quality. In the present research, in order to assess initially the data, it was necessary to scrutinize available data sources in unite of quality control of company at first. Initially, the quality and the accessibility of available data was scrutinized, before collection and choice of data in order to launch the preprocessing applicative.

Preparation of data

The present stage consists of all activities which is applied from primitive vain data for construction of final data collection (the data which is prepared for model). As much as the quality of the preparation is better, the model is better too. The assignment of data preparation is carried out in several stages; and there is not any pre determination sequence. This assignment consists of the choice of tables, records, features, and also conversion and purgation of data for model.

Model

In the present stage, it was chosen and applied all type of models. Totally, there is several methods for one data analysis problem. Some of the methods require a special format. Therefore, it frequency is necessary to refer the preparation stage. In the present study in order to apply the methods was utilized the CLEMENTINE 12 program. In order to model the data, each algorithm was categorized separately; and the algorithm which had the most accuracy was assigned as the model and exploitation of knowledge. The purpose of it is to exploit of knowledge according to data and chosen algorithm.

Assessment of results

In this stage of the research, a model was structured which has high quality according to data analysis point of view. Before applying the model, it should be totally assessed and applied steps reviewed in order to correspond with the aim of the trading.

Application of the model

In this step, the model does not mean the end of the research; even if the aim of the model is to promote the data knowledge, the achieved knowledge is required an organization again. It should be presented in the way that the followers can use it. According to the requirements of the work, the stage of application, can be simplicity as making a report, or complicated as the application of repeated process of data analysis.

RESULTS AND FINDINGS OF THE RESEARCH

The research sequence was separated in five parts: making data sources, description of data, analyze the model, assessment of the model, and application of model results; the details of sequences have been presented in the following.

Making data sources

At first sequence, the format of saved information in previous month in unit of quality control was scrutinized, and inscribed in this data source, in order to assess the operation of sequences, by design of a proper information bank. The information which were related to the body making section and with information field like the number of body, the number of action, time of production, defect or inexistence of defect, type of defect, the situation of assessment, and the others. At recent month, was inscribed thousands records of information.

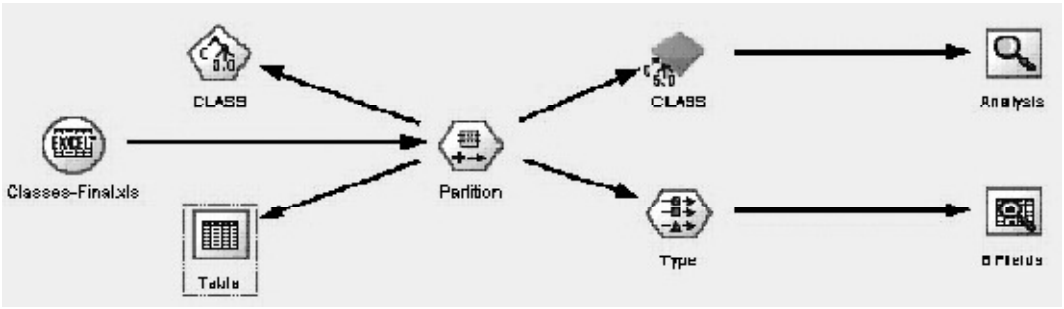

Figure 1. Modeling of algorithm of categorization of 5C

Table 1. The amount of accuracy of algorithm

Partition	Education		Experiment	
Correct categorization sample	709	90.9	303	92.1
Incorrect categorization sample	71	9.1	26	7.9
Total samples	780		329	

Therefore, data was collected and saved in a data source and then, data was comprehended and purged.

Description of data (summarization and making image of data)

In the present sequence, the information was computed according to body number, time and shift of work, the code of final controller, and the person's code of quality control unit, the code of defect and illustration of defect, an observation design, and illustration of observed level. The statistic society of the research consists of data related to about 3500 cases of bodies of Peugeot 206; which were produced at eleven month of 1394, and were scrutinized and controlled. Considered that data of statistic society was scrutinized according to their quality, and they were pure as much as possible, therefore, research models for final sample were carried out just from 1100 cases of body data of Peugeot 206; which were saved in a data source in excel program. The omitted records were related to the bodies which were not available in reactivity excel file according to the information about the structured body; or has faulty information.

Description of the model

In the present section, data was modeled, and required analysis was presented. At first sequence, in order to increase the accuracy and quality of information, and the special illogical cases and such cases will eliminated from collection of under process data by scrutiny and determination according to definition of comprehensive rules. The applied algorithm in the present study is decision tree C5; and the stream of model in CLEMENTINE 12 is as follows. The information of unit of Reactivity System was elided with unit of Shop Body. From two level of information and three model of decision tree C5, Logistic analyze, Chaid analyze, C&R Tree analyze were utilized in data analysis, which are in the indexes. (Figure 1)

The amount of accuracy of algorithm is presented in Table 1, according to correct categorization of data and according to the amount of repetition of experiment. As it is presented in the table 1, for instruction data from total number of 1100 records, 709 (90.9%) of data were categorized correctly; and 71(9.1%) records were categorized incorrectly. Regarding experimented data, from the total number of 329 records, 303 (92.1) records were categorized correctly, and 26 (7.9%) records categorized incorrectly.

Assessment of validity

Whereas the validity of the presented method should be assessed in every research, in order to assess the validity and accuracy of the method, it was utilized the division of sample in two collections of instructional and experimental data. The adjustment of validity is scrutinized according to the result of data; and experimental data was entered the algorithm as observer,

Table 2. The accuracy of result of the model

C&R	Logistic	C5	Algorithm
99.1	99.8	99.8	Accuracy

and the results assessed the adjustment of accuracy. In the present research, from the total number of data, 90% was chosen from instructional data, and 10% was chosen from experimental data. The adjustment of validity and accuracy of the model is depended on the accuracy of categorization or demodulation of experimental data. In order to assess the application of presented model of decision tree C5, The accuracy of result of the model was compared with the other categorized algorithm like, neural network, decision tree C&R, Logistic Regression; which the following table represents the results. (Table 2)

As the above table shows the decision tree C5 and logistic Regression have the same and higher (99.8) accuracy than C&R method. As regards the number of leaves and size of this tree is more than the other trees, therefore the best algorithm was chosen.

Application of results

The achieved information was accessed for the quality control unit, and utilized as a source in order to distribute quality documentation for body section in direction of accuracy in inspection and determination of proper person in order to decline the existing faults in reactivity unit and decline of dissatisfaction of users. The following table is a part of output of observable program, which the data was divided into two parts, learning and experiment; and the results were the same.

CONCLUSION AND SUGGESTIONS

Today companies and industrial organizations confront a numerous data and information, which was constantly collected from routine activities for the aims like control and guaranty of production quality, schematization, and control of operation of production lines. Today, there is spacious information sources in these organizations from corresponsive information with generative and supportive process which finding of patterns and hidden knowledge from data can facilitate the determiners of industry arena in order to increase the quality and productivity of their productions. Whereas the decision tree C5 has the higher accuracy in contrast the other algorithm of categorization of data analysis, therefore the exploited knowledge from this tree, is the most trustable knowledge from data which was scrutinized in the present study; and it can be a source in order to exploit the rules of data analysis. In the present study, a case was scrutinized for application of data analysis techniques about quality control process and anticipation of accuracy of quality operation of persons in body section and preventing of repetition of fault from this unit in reactivity section. The result of the study showed that in 85% occasions, the anticipation is trustable. Naturally, if the records were inscribed more accurate in diverse production units and also access the more comprehensive data can utilize and anticipate mostly in order to decline dissatisfactions of quality of Iran Automotive productions. Also, there is a numerous content of information in many generative units and the other sections of this company; which is not utilized. For further researches, it seems to be interesting to collect data from diverse unit of montage, and in more spacious and greater dimensions.

REFERENCES

Azar A (2011). the design of Power and Humanity according to data analysis. Technical information management. (4):22-23.

Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R (2000). *CRISP-DM Step-Data Mining Guide*. SPSS Inc.

CRISPMWP.

Fayyad U, Piatetsky G, Smyth P (1996). “From Data Mining to Knowledge Discovery in Databases”.

Han J, Kamber M (2001). "Data Mining: Concepts and Techniques", San Diego Academic Press.

Hand D, Mannila H, Smyth P (2000). “Principles of Data Mining”, the MIT Press

Harding JA, et.al. (2006). “Data Mining in Manufacturing: A Review”, Journal of Manufacturing Science and Engineering, Vol. 128, Nov.

Last M, Kandel A (2001). “Data Mining for Process and Quality Control in the Semiconductor Industry”, Data Mining for Design and Manufacturing: Methods and Applications, D. Braha, ed., Kluwer Academic, pp. 207–234

Mahmoodi K, Rostami H, Moradi E (20014). The review of data analysis. Sharif University. Sahel Convention. First of June.

Neaga EI, Harding JA (2005). “An Enterprise Modelling and Integration Framework Based on Knowledge Discovery and Data Mining”, Int. J. Prod. Res., pp. 1089–1108.

Romanowski CJ, Nagi R (2001). “Analysing Maintenance Data Using Data Mining Methods,” Data Mining for Design and Manufacture: Methods and Applications, D. Braha, Kluwer Academic, Dordrecht, pp. 235–254.