

# A Review of Business Intelligence Techniques for Mild Steel Defect Diagnosis

Veena Jokhakar  
M.Sc. (I.T.) Programme  
Veer Narmad South Gujarat University  
Surat

S.V. Patel, Ph.D.  
Computer Science Department  
Veer Narmad South Gujarat University  
Surat

## ABSTRACT

In this competitive era, manufacturing companies have to focus on the quality of the produced products. The quality of the product produced is affected by many influential parameters during the process. The product once produced with a lower quality then usually ends up with incurring loss in certain terms to the company. Hence, it is extremely important to know the defect causing parameters and perform defect diagnosis. Various techniques like SPC-SQC, Six-Sigma and Kaizen have been used for quality analysis. But since last few years machine learning and data mining is being used for analysis due to advancement in the field and its advantages. This paper conducts an analytical survey of various business intelligence techniques used in for defect diagnosis. The paper concludes with the analytical results as random forest performs the best in terms of performance compared to other techniques and shows the future research scope in this area. Moreover, we find that random forest has not been introduced yet in steel defect diagnosis.

## General Terms

Statistics, data mining, neural network, genetic algorithm, and ensemble approach.

## Keywords

Ensemble approach, random forest, steel defect diagnosis

## 1. INTRODUCTION AND BACKGROUND

Survival of the fittest is the rule for surviving successfully in the world of competition. For the same the management needs to be on toes and take quick, correct and firm decisions frequently for collective growth. Business intelligence (BI) is a technology-driven process for analyzing data and useful information to help business managers and end users make precise decisions. BI encompasses a variety of tools, applications and methodologies that enable organizations to collect data from various internal and external sources, prepare and perform analysis to create reports, dashboards and data visualizations. BI programs include forms of advanced analytics, such as data mining, predictive analytics, text mining, statistical analysis and big data analytics.

The products manufactured by manufacturing companies' products may vary in quality than the expected quality in terms of many parameters. This expected quality may be customer defined. And if the quality produced deviates from what was expected by customer, then the product may be sold at a lower rate to the customer or may sold to some other customer who may accept the product. This incurs loss to the company in huge amount in terms of money and time.

Statistical process and quality control has been used for quality analysis of products during the manufacturing process like x-bar and r-chart, as well as post analysis of defects like p-charts etc. However, manufacturing process are usually so complex that traditional statistical techniques or data management tools are

not sufficient to extract learning patterns and knowledge for the quality improvement.

In order manage this problem, data mining approaches are found useful to extract knowledge from the data generated in the process. Data mining predicts future trends and behavior which makes businesses upbeat, knowledge-driven decisions [1]. Data mining algorithms can be divided as classification, clustering, regression and association models.

Classification models classify data and properly calculate the value of each class variables [2,3]. There are different types of classification methods such as bayes functions, neural networks, support vector machines (SVM), decision trees and fuzzy rules.

Clustering is a unsupervised method of machine learning application that clusters the dataset in groups with certain similarity in features. Examples of clustering algorithms are K-means, Cobweb, farthest first etc. Amongst them K-means is most widely used and is simple.

Regression models like logistic regression, linear regression, and multivariate non –linear regression are used for modeling a random variable, B as linear function of another variable A. Where A is the predictor variable and B is the response variable. Mathematically  $B = a + bA$ . Where a, b are regression coefficients that control the growth pattern denoting Y intercept and slope of a line. Multiple regressions are used for more than one predictor variables.

Association rules mining models are used for discovering the relationships and patterns in large databases. Two statistical measures known as Support and Confidence are used to control the process of association. Where support shows the significance of the rule and confidence shows the degree of certainty that the rule is satisfied in the data set. It takes the mathematical form as  $X \rightarrow Y$ , where X and Y are data items. This is interpreted as the presence of X implies presence of Y also. There are many association rule-mining algorithms alike APRIORI and CARMA.

In recent years data mining began to be applied to quality diagnosis and quality improvement in complicated manufacturing process, such as semiconductor manufacturing and steel making [4]. Many recent researches in fault detection and prediction have used data mining in their work. This paper does an analytical survey of the various recent approaches applied for fault diagnosis and then we particularly discuss tree ensembles, that has recently gained interest in research field since few year in various fields and stress on its exploration in fault diagnosis.

The paper has been organized as follows: Section II presents detail survey of dominant models used defect prediction especially in steel industry, III Discusses the analytical review and Section IV Concludes the paper.

## **2. DOMINANT MODELS FOR DEFECT DIAGNOSIS**

Various techniques have been applied for fault diagnosis using BI in industries. The approaches used by various researchers in their respective work.

### **2.1 Statistical Approach**

Multivariate statistical methods proved to be easy to implement and satisfactory for basic industrial tasks, however many of them are linear and not give accurate information. So to improve prognostic system nonlinear techniques should be taken into consideration [4]. Linear PCA can handle high dimensional and correlated process variables, provides a natural solution to the errors-in-variables problem and includes disturbance decoupling, but it is linear therefore a lots of research was invested to nonlinearity and fitting to the nonlinear processes. According to their survey prognostics of complex engineered systems remains an area in which much more research is needed. Hence, Artificial intelligence and soft computing methods can offer great results, especially if combined into hybrid platforms.

[4] Also discuss that Statistical process control and design of experiment approaches did not provide conclusive results. The DM answers this knowledge discovery concern as it is defined as a technique to extract predictive information and knowledge from databases [5].

[6] Used logistic regression, which is one of the traditional techniques used to determine the most influential process factors on the target process variable. They discusses the application of regression model in defect cause analysis of measured on 36 input variables on 809 defective items produced. In this study, they have tried to develop a regression model which relates defect types to input factors. As the response variable was of nominal type they have applied multinomial logistic regression approach using Clementine 10.1, the data mining software of SPSS. Although the fitted model was statistically significant (p-values for Pearson and Deviance was 1.0; p-value for G was 0.0), none of the parameter estimates are found to be significant.

Experts in [7] state that as a key aspect of quality Control and diagnosing, this root cause identification involves searching for systematic faults that explain the observed variability behavior by incorporating process knowledge.

They propose a statistical tool for diagnosing the quality of solder pastes, the proposed MLPCA based regression coefficients clustering algorithm.

Disadvantages of this approach is that using statistical techniques and interpreting their results is difficult and requires a considerable amount of knowledge of statistics that non statisticians find difficult to resolve and understand.

### **2.2 Decision Tree Machine Learning Approach**

In [8] they also have applied decision tree which is a simple tree-shaped structure where each internal node represents a test on one attribute, arcs show the results of a test and leaf nodes reflect classes. They are easy to understand and interpret. They can be applied to classify both numeric and categorical data. Because of these advantages decision trees are used extensively for prediction and classification purpose.

[8] Have used C5.0 algorithm, which is an improved version of C4.5. They have used Clementine 10.1 to implement this algorithm. After building the tree, global pruning with 75% pruning severity was performed to avoid over fitting. Leaf nodes were allowed with minimum 5 records. Estimated accuracy for

the final model was found to be 92.15% for the training set. The model classified 91.93% of the testing data correctly. The decision tree model found nine process variables to be influential on the response, defect types, and it also extracts ten rules associated with these significant input variables.[8] Also show that logistic regression although used for defect analysis, gave unsatisfactory results. Instead CART I results in 64 % accuracy and CART II resulted in 92% accuracy.

In [9] the author has proposed a knowledge based continuous quality improvement in manufacturing quality. They state that DMAIC(Define-Measure-Analyse-Improve-Control) is a problem driven approach. They have proposed a different model from DMAIC that is goal-driven approach. They have collected 1000 records randomly from the process. Where each records consists of 4 factors. The records were classified on defect rates. The records with defect rate > 3.0% were set as H and others as L. They have used Decision tree C5.0 for the analysis. They generated 5 rules. They have used these rules for identifying parameter optimization.

[10] Established an algorithmic decision tree, for the prediction of cracks evolution during hot rolling. They evaluated prediction of V-shaped and U-shaped cracks in steel slab during hot rolling using C4.5 program.

Fahmi Arif, et.al.[11] have used a combination of PCA and ID3 algorithm in multistage with 1115 instances with 22 variables to realize more faultless manufacturing. Each PCA model has its own weight that is defined by weighted clustering technique. [12] Shows the use of decision trees to create a control mechanism for regulation of the Hot metal temperature (HMT), a state of blast furnace for examining the feasibility of structuring a control mechanism for the HMT of pig iron utilizing available decision tree and rule-based predictor formulation technology.

Advantages of this approach is that it is very easy to interpret and visualize. It works with numeric and categorical data.

### **2.3 Neural Network Machine Learning Approach**

Jarno J. Haapamali, et.al. [13] Have used data mining methods in hot steel rolling for scale defect prediction. The data they gather were average values and process measured with different frequencies. Origin of scale defects has been a topic of interest for many research projects, but it is still hard to find literature on the modeling of defects [13]. They have gathered 1326 steel strips and 50 variables from diagnostics measurements and 127 averages values. They used SOM neural networks for this task, as it is capable to visualize non-linear dependencies and has been used in various industrial applications. The mean accuracy of the system was probably lower than 90%. [14] Also shows the applicability of regression and neural network in defect detection at Tata steel.

[15] Presents the preliminary results of a data mining study of a production line involving hundreds of variables related to mechanical, chemical, electrical and magnetic processes involved in manufacturing coated glass. The study was performed using two nonlinear, nonparametric approaches, namely neural network and CART, to model the relationship between the qualities of the coating and machine readings.[16] Observers good results when using multi-layered feed-forward artificial neural network (ANN) models to predict the silicon content of hot metal from a blast furnace. As time-dependencies (time lags) between each of the various inputs used and the output (silicon content) were found to be significant, each input was lagged from the output by a set amount. [17] examines the ability of feed-forward neural networks to predict the hot metal temperature of pig iron based

on 39 input parameters that reflect the current and past conditions of the blast furnace.

Authors in [18] propose a structured model based on neural network with radial basis function for defect component analysis along with PCS.

Advantage of this approach is that this can be applied to both linear and non linear data. uses less statistical training, has the ability to detect all possible interactions between predictor variables.

Disadvantages of the approach are that it is difficult to interpret, "black box" nature, greater computational burden and is prone to overfitting.

## **2.4 Support Vector Machine Learning Approach**

Authors in [19] have shown a comparison between various mining algorithm for fault detection. They conclude support vector machines SVM have the best processing time and also overall high accuracy. Then they applied C5 model to generate predictions rules.[20] Also propose a (SVM) based model which integrates a dimension reduction scheme to analyze the failures of turbines in thermal power facilities. They have taken case from a thermal power plant to evaluate the effectiveness of the proposed SVM based model. They conclude that that SVM outperforms linear discriminant analysis (LDA) and back-propagation neural networks (BPN) in classification performance.

In [21] authors have proposed a new approach for fault detection and diagnosis based on One-Class Support Vector Machines (1-class SVM) has been proposed. The approach is based on a non-linear distance metric measured in a feature space.

Disadvantages of this technique are its high algorithmic complexity and extensive memory requirements in large-scale tasks.

## **2.5 Association Mining**

Authors in [9] also stated that during manufacturing the process variable undergo variation. For that purpose they have used SPC to identify the variations and then used association rule for diagnosis of cause of the problem. Here apriori algorithm minimum antecedent support was set to 0.65% and confidence was set to 80%. They considered 62,592 products, 16,905 products failed the test. They concluded that 95% of the products that failed in the test were due to assembly line 4.

[22, 23] Mining association rules is particularly useful for discovering relationships among items from large databases. Authors of [24] have used Association rules, Decision Tree and neural networks for detection of steel defects on surface. After performing preprocessing continued with the selection of the most meaningful variables, and the number of variables was reduced from 186 to 36 at the modeling phase. The study also tried using these algorithms to reduce product defects with Pits & Blister defect.

They show that the accuracy percentage of each model out performs other depending on the carbon content and other elements.

This has been shown in table 1.[25] Also shows Root-cause Machine Identifier (RMI) method using the technique of association rule mining to solve the problem of defect detection efficiently and effectively. This analysis was defect case based.

## **2.6 Genetic Algorithm Machine Learning Approach**

For prediction of scale defect prediction [26], shows the Genetic algorithm (GA) based method. They have used SOM for identifying the most promising variables and then used these variables in GA for the prediction. Average error was 0,0957[1/m<sup>2</sup>]. [27] uses genetic algorithm to create a controller for the HMT. Given current conditions (specified by the current HMT, which is referred to as the "operating point," and current values of the input variables), the solution should determine what changes need to be made to each variable in order to achieve a desired HMT at a desired time (some hours into the future).

## **2.7 Rough Set Theory Approach**

Error rates at a manufacturing process are used as input to identify knowledge for further assistance to engineers; authors of [28] comment. They have presented presents a new heuristic algorithm, called extended rough set theory, for reduct selection in rough set theory (RST) applications. They also show a comparison of SPC-SQC control charts and regression analysis for the same.

## **2.8 Ensemble based Approach**

In recent years, a number of data mining approaches for modeling data containing non linear and other complex dependencies have appeared in literature[29].[30] shows the effective use of Random forest in non-linear insurance claim data. They investigated BART, S-PLUS, TREENET, logistic regression, Random Forest and Insightful Miner Tree procedures. [29] Shows the results of decision tree ,tree forest and tree boost approaches. Tree boost and Tree forest both create ensemble of trees. The difference is tee boost creates it in sequence and forest creates them in parallel. Both the trees produce high accuracy, but surely results better than single tree. Previous research [31] has shown that an ensemble is often more accurate than any of the single classifiers.

[32] Evaluated 23 datasets with neural networks, decision trees and ensemble approaches. And they concluded that bagging is always more accurate than any other classifier.[33] Made use of disjoint partitions of training datasets to learn the base decision trees, and ranking of training bootstrap samples on the basis of diversity. The approach lead to efficient learning of Random

Forest classifier. [34] states that for a small number of variables, the ANN classification was competitive, but as the number of variables was increased, the boosting results proved more efficient and superior to the ANN technique. [35] Found that random forest feature extraction showed comparable fault diagnosis performance for The Tennessee Eastman process, better fault identification performance for the simple nonlinear system, and better fault detection performance for the calcium carbide process; as compared to principal component analysis. [36]have presented a taxonomy of Random Forest algorithm and performed analysis of various algorithms / techniques based on Random Forest algorithm.[37] proposed a random forest based research that obtained 99.7% classification accuracy in diagnosis and prognosticating breast cancer with ML and state that random forest results better than other ML algorithms done in past works.

The effectiveness of random forests on fault diagnosis of power electronic circuit experimented by determining optimal configurations of random forest in [38].Experimental results show that the method is feasible and effective. A total of 1760 data of all 22 classes were selected for fault testing and accuracy of 98.92% was seen.

[39] Applied two different types of randomness, inserted into the individual trees of a random forest. They evaluated results obtained by other classical Machine Learning algorithms, such as neural networks, Classification and Regression Trees (CART), Naive Bayes, and K-nearest neighbor (KNN). They used a dataset of Gas turbine for fault diagnosis with 864 instances described by 27 distinct attributes. Best results were obtained by using 500 trees and 6 features. This research achieved performance 97% precision and recall.

Authors in [40] state that random forest has achieved good results when being applied to medicine, biology machine learning and other areas but however they apply this method to machinery fault diagnosis. They experiment RFA on ship intelligent fault diagnosis of chain box. They concluded that the error rates with 3 dimensions were less than 4 dimension RF creation. They have used 220 samples with 16 dimensions.

A comprehensive table of survey is shown in Table1. This shows that though the existing MLA have good performance like decision trees, neural networks and genetic algorithms, still random forest outperforms these MLAs. This seems to be an emerging MLA and is yet to be explored in many fields and provides a scope for further research based on its applicability in various industries like manufacturing, stock market, medicine etc, applicability in various fields such as stream data mining and performance up gradation.

### 3. ANALYTICAL RESULTS

We here present the analytical survey in the form of table in Table 1. “Ref. No” column is to the reference number referred, “method used” shows the approach followed in the respective work, “problem addressed” describes the problem on which the approach was applied, “input variables and instances” shows the amount of data used for addressing the issue, “method used” is the algorithm used, “statistically significant, error rates, accuracy” show any statistically relevant measure of performance achieved or error rates obtained or accuracy obtained, which as addressed as S,E, and A respectively and “comments” describes the final conclusion of the applied approach. A symbol “\*” shows that, no value is present to assign the respective column.

We find that lots of work has been done in fault diagnosis in manufacturing industries. We studied that experts though have applied statistical approaches in many work but researchers emphasis that application on statistics only is not capable of exploring patterns and knowledge from the data, hence the application of data mining is highly recommended. Many experts have used different machine learning methodologies in their work, and have presented the performance achieved. We find that the performance of ensemble techniques particularly tree ensembles outperform other existing techniques. Tree ensembles are extended decision trees that generate hundreds or thousands of trees as optimized. These trees are created one by one using sampled data randomly, keeping 1/3 for out of bag error calculation. They can deal with “small n large p”-problems, high-order interactions and correlated predictor variables. They also are very stable and almost perfect “out of the box” classifiers. We further find that this powerful machine learning algorithm has been to various fields like genetics, bioscience, neuro science etc, and some work was found in stock market and software engineering. But we do not find any work done with this in steel industry.

**Table 1: Showing the comprehensive survey of the studies (those whose results are available in terms of accuracy)**

Ref. No.	Model Used	Problem addressed	Input Variables And Instances	Method Used	Statistically Significant(S),Error Rates(E),Accuracy(A)	Comments
6	Statistical Approach	<b>Defect cause analysis</b>	36,806	Logistic Regression	p-vale for Pearson and deviance = 1.0  And for G = 0.0 (S)	Unsatisfactory Results
7		<b>Route cause identification of fault</b>	*	MLPCA based regression coefficients clustering algorithm	*	Case based results
8		<b>Defect analysis</b>	*	Logistic Regression	*	Unsatisfactory results

10	Decision Tree Approach	<b>prediction of cracks evolution</b>	5,*	C4.5	*	Evaluated U,V shaped cracks.
11		<b>Faultless manufacturing</b>	22,115	PCA and ID3 in multistage	90(A)	Probability of misclassification of negative class is high with combination of PCA+ID3 than other techniques.
6		<b>Defect cause analysis</b>	36,806	C5.0	91.93(A)	Proved to be better than Logistic Regression
8		<b>Defect analysis</b>	*	CART I CART II	64(A) 92(A)	CART II was better than CART I and Logistic Regression
9		<b>Continuous quality improvement</b>	*,1000	C5.0	*	Generated 5 Rules and used these for Parameter optimization

24		<b>Prediction of defects on steel surface</b>	*,DS1-2300 DS2-350	Decision tree	83(A)	For low carbon products(0.0% -0.15%)
----	--	---	-----------------------	---------------	-------	--------------------------------------

35	Ensemble Approach	<b>Improve the performance of fault diagnosis</b>	41,960	Random Forest Tree Boost	98(A)	Tree boost was comparatively accurate than single tree
37		<b>Breast cancer diagnosis and prognosis</b>	30,151	Random Forest	99.82(A)	*
38		<b>fault diagnosis of power electronic circuit</b>	1760 of 22 cases	Random Forest	98.92(A)	RF is recommended because of fast execution speed, the characteristic of tree classifier, and high process-speed
39	Ensemble Approach	<b>Gas turbine for fault diagnosis</b>	864 instances with 27 attributes	Random Forest	97(A)	Found Random forest to outperform among other ML algorithms like neural networks, Classification and Regression Trees (CART), Naive Bayes, and K-nearest neighbor (KNN)

40		<b>machinery fault diagnosis on chain box of ships</b>	220 samples with 16 dimensions	Random Forest	9.52(E)	3 dimensions had less error rates than 4 dimensions in each node
24	Neural Network	<b>Prediction of defects on steel surface</b>	*,DS1-2300,DS2-350	Neural Network	76(A)	products with molybdenum
13		<b>Scale defect prediction</b>	50 + 127 avg. val.,1326	SOM neural networks	<90(A)	Visualization of the rolling and defect modeling
9	Association Mining	<b>Identify variations and diagnose cause</b>	*,62592	SPC for variations identification and Association Rules Mining	*	95% of the products failed due to assembly line 4
24		<b>Prediction of Pits and Blister Defect on steel surface</b>	*,DS1-2300,DS2-350	Association rules	64(A)	For low carbon products(0.15% - 0.25%)
19	SVM	<b>Fault detection</b>	8,*	SVM and others	90(A)	SVM proved to be better than other

13	Genetic algorithm	<b>Scale defect prediction</b>	50 + 127 avg. val,1326	Genetic algorithm	0.0640(E)	Performance not as good as neural network.
----	-------------------	--------------------------------	------------------------	-------------------	-----------	--

#### 4. CONCLUSION

The intension of this paper is to present a survey of current work related to defect prediction done and the future research directions in the field. As a result of the survey we conclude that statistics, decision tree and neural network are wide used ML algorithms of the purpose of fault diagnosis. Some recent works in Random Forest show that they give better performance than ANN and Single Decision Tree, as they are an ensemble of trees and hence is more accurate, but it is time consuming compared to other individual classification techniques. We also see that Random forest show better results than other machine learning algorithm and has been used widely in medicine and biotechnology area in recent years, but has not been evaluated in steel industry for fault diagnosis. This also show that there are various work with combination of two approaches like statistics with neural networks, but no work has been done with application of statistics and random forest. This analysis will serve as a guideline for pursuing future research related to fault diagnosis in steel.

#### 5. REFERENCES

- [1] Dharminder Kumar, Suman, "Performance Analysis Of Various Data Mining Algorithms: A Review", *International Journal Of Computer Applications (0975 – 8887)*, Volume 32– No.6, October 2011
- [2] Daniel Grossman and Pedro Domingos, "Learning Bayesian Network Classifiers by Maximizing Conditional Likelihood", 21st International Conference on Machine Learning, Banff, Canada, 2004
- [3] A Review of Clustering and Classification Techniques in Data Mining, Samir Kumar Sarangi, Dr. Vivek Jaglan, Yajnaseni Dash, In *ternational Journal of Engineering, Business and Enterprise Applications (IJEBA)*, ISSN (Print): 2279-0020 ISSN (Online): 2279-0039
- [4] Bratina Božidar And Boris Tovornik, *Multivariate Statistical Methods For Industrial Process Prognostics, Cybernetics And Informatics*, International Conference February 10 - 13, 2010
- [5] Kesheng Wang, *Applying Data Mining To Manufacturing: The Nature And Implications*, *Journal Of Intelligent Manufacturing*, 2007
- [6] B. Bakır, Bakır, I. Batmaz, F. A. Güntürkün, I. A. Ipeççi, G. Köksal, And N. E. Özdemirel , *Defect Cause Modeling With Decision Tree And Regression Analysis*, *World Academy Of Science, Engineering And Technology International Journal Of Mechanical, Aerospace, Industrial And Mechatronics Engineering Vol:2 No:12*, 2008
- [7] Feng Zhang ; Fairchild Semicond., South Portland, ME ; Timwah Luk, *A Data Mining Algorithm for Monitoring PCB Assembly Quality* , *Electronics Packaging Manufacturing, IEEE Transactions on (Volume:30 , Issue: 4 )*, ISSN : 1521-334X, Oct. 2007
- [8] Berna Bakır, *Defect Cause Modeling With Decision Tree And Regression Analysis: A Case Study In Casting Industry*, A Thesis Submitted To The Graduate School Of Informatics Of Middle East Technical University, May 2007
- [9] Shu-Gauge He, Zhen He, G. Alan Wang And Lili, *Quality Improvement Using Data Mining In Manufacturing Processes*, *Data Mining And Knowledge Discovery In Real Life Applications*, Isbn 978-3-902613-53-0, Pp.438, Feb 2009
- [10] Martin Vlado,1 Róbert Bidulský,2; Lucia Gulová,1 Kristína Machová,1 Jana Bidulská,1 Ján Valíček3 And Ján Sasl, *The Production Of Cracks Evolution In Continuously Cast Steel Slab, High Temp. Mater. Proc.*, Vol. 1–2 (2011), Pp. 105–111, Copyright © 2011 De Gruyter. Doi 10.1515/Htmp.2011.014
- [11] Fahmi Arif, Nanna Suryana, Burairah Hussin, *A Data Mining Approach For Developing Quality Prediction Model In Multi-Stage Manufacturing*, © 2013 By Ijca Journal, Volume 69 - Number 22
- [12] Mu'taz M. Qubbaj, Supervised By Dr. Amay Gupta, *Using Decision Trees To Create A Control Mechanism For Regulation Of The Hot Metal Temperature Of The "G" Blast Furnace At Steelcorp*, 2000.
- [13] Jarno J. Haapamaki, Satu M. Tammiminen And Juha J. Roning, *Data Mining Methods In Hot Steel Rolling For Scale Defect Prediction*, , *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications*, 2005
- [14] Shankar Kumar Choudhari, Sunil Kumar, Vinit K Mathur, *Application Of Datamining Technology At Tata Steel*, 2004
- [15] M. Li. S. . Sethi, J. Luciw, K. Wangner, *Mining Production Data With Neural Network & Cart*, In *Conf. Rec. IEEE Int. Conf. Data Mining*, 2003.
- [16] Bulsari, Abhay And Saxen, Henrik. "Classification Of Blast Furnace Probe Temperatures Using Neural Networks." *Steel Research*. Vol. 66. 1995
- [17] Sujit Kumar Bag, *ANN BASED PREDICTION OF BLAST FURNACE PARAMETERS* , *Journal - The Institution of Engineers, Malaysia (Vol. 68, No. 1, March 2007)*
- [18] D. Flynn, J. Ritchie And M. Cregan, *Data Mining Techniques Applied To power Plant Performance Monitoring*, *Processdings of the 16th IFAC Worl Congress*, 2005 , Volume 16 , Part 1, pp.1636-1636
- [19] Golriz Amooee, Behrouz Minaei-Bidgoli, Malihe Bagheri-Devnabi, *A Comparison Between Data Mining Production Algorithms For Fault Detection*, *Ijcsi* , Vol.8, Issue 6, No 3, Nov 2011, Issn :1694-0814
- [20] Kai-Ying Chen, Long-Sheng Chen, Mu-Chen Chen, Chia-Lung Lee , *Using Svm Based Method For Equipment Fault Detection In A Thermal Power Plant*, *Computers In Industry* 62, 42–50, 2011 *Journal Homepage: Www.Elsevier.Com/Locate/Compind*
- [21] Sankar Mahadevan And Sirish L. Shah, *Fault Detection And Diagonosis In Process Data Usinh Support Vector Machines*, *Jounal Of Process Control* 19, 1627-1639, 2009
- [22] Yu-Chiang Li, Jieh-Shan Yeh, Chin-Chen Chang, "Efficient Algorithms for Mining Shared-Frequent Itemsets", In *Proceedings of the 11th World Congress Of Intl. Fuzzy Systems Association*, 2005.
- [23] V. Umarani And Dr. M. Punithavalli, *A Study On Effective Mining Of Association Rules From Huge Databases*, *Ijcsr*, Vol. 1 Issue 1, 2010
- [24] Sayed Mehran Sharafi, Hamid Reza Esmaeily, *Applying Data Mining Methods To Predict Defects On Steel Surface*, *Jounal Of Theoretical And Applied Information Technology*, ©2005-2010

- [25] Wei-Chou Chen, Shian-Shyong Tseng, Ching-Yao Wang\*, A novel manufacturing defect detection method using association rule mining techniques, ELSEVIER,Expert Systems with Applications 29,807–815, 2005
- [26] Jarno Haapamaki And Juha Roning,Genetic Algorithms In Hot Steel Rolling For Scale Defect Prediction,World Academy Of Science ,Engineering And Technology 5, 2007
- [27] By Danny Lai, Supervised By Dr. Amar Gupta, Using Genetic Algorithms As A Controller For Hot Metal Temperature In Blast Furnace Processes, 2000
- [28] Tzu-Liang (Bill) Tseng, M.C. Jothishankar, Tong (Teresa)Wu, Quality Control Problem In Printed Circuit Board Manufacturing ,An Extended Rough Set Theory Approach, Journal Of Manufacturing Systems, Vol. 23/No. 1,2004
- [29] Tarun chopra and Jayashi Vajpi, Fault Diagnosis Benchmark Process Control System Using Stochastic Gradient Boosted Decision Trees, Ijsce,Issn:2231-2307,Vol-1,Issue-3,Jul 2011
- [30] Richard Derrig, Ph.D. And Louise Francis, Distinguishing The Forest From The Trees: A Comparison Of Tree Based Data Mining Methods, Casualty Actuarial Society Forum, Winter 2006
- [31] Krogh A, Vedelsby J, Neural Network Ensembles, Cross Validation, and Active Learning, Advances in Neural Information Processing Systems Vol 7, MIT Press , 231-238, 1995
- [32] Opitz D, Maclin R, Popular Ensemble Methods: An Empirical Study, Journal of Artificial Intelligence 11, 169-198, 1999
- [33] Vrushali Y Kulkarni,Pradeep K Sinha, Effective Learning and Classification using Random Forest Algorithm , International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 11, May 2014, ISSN: 2277-3754, ISO 9001:2008 Certified
- [34] Byron P. Roa, Hai-Jun Yanga,\_, Ji Zhub, Yong Liuc, Ion Stancuc, Gordon Mcgregord, Boosted Decision Trees As An Alternative To Artificial Neural Networks For Particle Identification, Nuclear Instruments And Methods In Physics Research A 543 (2005) 577–584,Science Direct, Available Online 25 January 2005
- [35] Lidia Auret , Process Monitoring and Fault Diagnosis using Random Forests , Dissertation presented for the Degree of Doctor Of Philosophy (Extractive Metallurgical Engineering) , December 2010
- [36] Vrishali Y Kulkarni,Dr. Pradeep K. Sinha, Random Forest Classifiers:A Survey and Future, International Journal of Advanced Computing, ISSN:2051-0845, Vol.36, Issue.1, April 2013
- [37] Cuong Nguyen,Yong Wand,Ha Nam Nguyen,Random forest classifier combined with feature selection for breast cancer disgnosisi and prognostic,Journal Biomedical Science and engineering ,6,551-560,May 2013
- [38] Jin-Ding Cai, Ren-Wu Yan, Fault Diagnosis of Power Electronic Circuit Based on Random Forests Algorithm, 2009 Fifth International Conference on Natural Computation,978-0-7695-3736-8/09 © 2009 IEEE
- [39] Manolis Maragoudakis, Euripides Loukis and Panagiotis-Prodromos Pantelides, Random Forests Identification of Gas Turbine Faults, 19th International Conference on Systems Engineering, 978-0-7695-3331-5/08 © 2008 IEEE
- [40] Zhiyuan Yang, Qinming Tan ,The Application of Random Forest and Morphology Analysis to Fault Diagnosis on the Chain box of ships, Third International Symposium on Intelligent Information Technology and Security Informatics, 978-0-7695-4020-7/10 © 2010 IEEE