

A SURVEY ON DATA MINING IN STEEL INDUSTRIES

S. Umeshini¹, C.PSumathi²

¹ResearchScholar, Department of Computer Science, SDNB Vaishnav College for Women, Chennai, India

²Associate Professor and Head, Department of Computer science, SDNB Vaishnav College for Women, Chennai India

ABSTRACT

In Industrial environments, huge amount of data is being generated which in turn collected in database and data warehouses from all involved areas such as planning, process design, materials, assembly, production, quality, process control, scheduling, fault detection, shutdown, customer relation management, and so on. Data Mining has become a useful tool for knowledge acquisition for industrial process of Iron and steel making. Due to the rapid growth in Data Mining, various industries started using data mining technology to search the hidden patterns, which might further be used to the system with the new knowledge which might design new models to enhance the production quality, productivity optimum cost and maintenance etc. The continuous improvement of all steel production process regarding the avoidance of quality deficiencies and the related improvement of production yield is an essential task of steel producer. Therefore, zero defect strategy is popular today and to maintain it several quality assurance techniques are used. The present report explains the methods of data mining and describes its application in the industrial environment and especially, in the steel industry.

KEYWORDS

Repository, Explanatory variables, Clusters, Dependent variables, Ensemble methods, Decision making, patterns.

1. INTRODUCTION

In most steel sectors, [1] manufacturing is extremely competitive and financial margins that differentiate between success and failure are very tight with established industries needed to compete, produce and sell at global level. To master trans-continental challenges, a company must achieve low cost production yet still maintain highly skilled, flexible and efficient workforces who could consistently design and produce high quality, low cost products. This can be achieved by using data mining techniques to improve decision making.

Data Mining can be generally said as a technique to find patterns (extraction) or interesting information in large amount of data [2]. This technique has been widely used in research areas like engineering, marketing, business, education and now especially in industries like Iron and Steel, Rubber etc. However, knowledge can take many forms and it is necessary to identify the kind of knowledge to be mined when testing the huge amount of data generated during manufacturing. Data Mining is the process of interesting patterns and knowledge from large amount of data. The data source can include databases, data warehouses, the web, other information repositories, data that streamed into the system automatically [3].

2.NINE LAWS OF DATA MINING [4]:

First Law: “Business goals Laws” -Business objectives are the origin of every data mining solution.

Second law: “Business knowledge Law” - Business knowledge is control to every step of the data mining solution.

Third law: “Data preparation Law” - Data preparation is more than every data mining process.

Fourth law: “NFL – DM” - The right model for a given application can only be discovered by experiment.

Fifth law: “Watkins Law” - There are always patterns

Sixth law: “Insight Law” - Data mining amplifies perception in the business domain.

Seventh law:“prediction Law” - Prediction increases information locally by generalization.

Eighth Law:“Value Law” - The value of data mining results is not determined by the accuracy (or) stability of predictive models.

Ninth Law:“Law of Change” - All patterns are subject to change.

3.DATA MINING STRUCTURE

The architecture of a typical data mining system [3] has the following major components (Han &Kamber)

- Database, Data warehouse (or) other information repository.
- Database (or) Data warehouse server
- Knowledge base
- Data mining engine
- Pattern Evolution module
- Graphics user interface

Fayyad (1996) stated that data mining algorithms consist of some specific mix of these components.

3.1. Sources of data:

We get lot of data through internet and other sources. The data available over internet may be in different types. When we have huge amount of data, then it is big data, which creates another challenge in every aspect. The sources of data may be internal or external. The data may be structured, unstructured and semi-structured(Figure1). Such data are analyzed by predefined tools and technique referred to as data mining. It creates trends and patterns in data.

The data may be summarized as follows [5]:

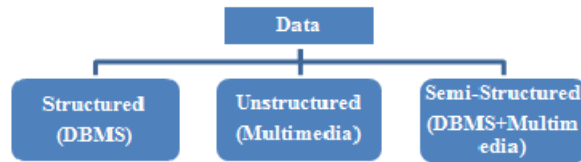


Figure1. Data Summarization

4. THE ITERATIVE PROCESS IN KDD

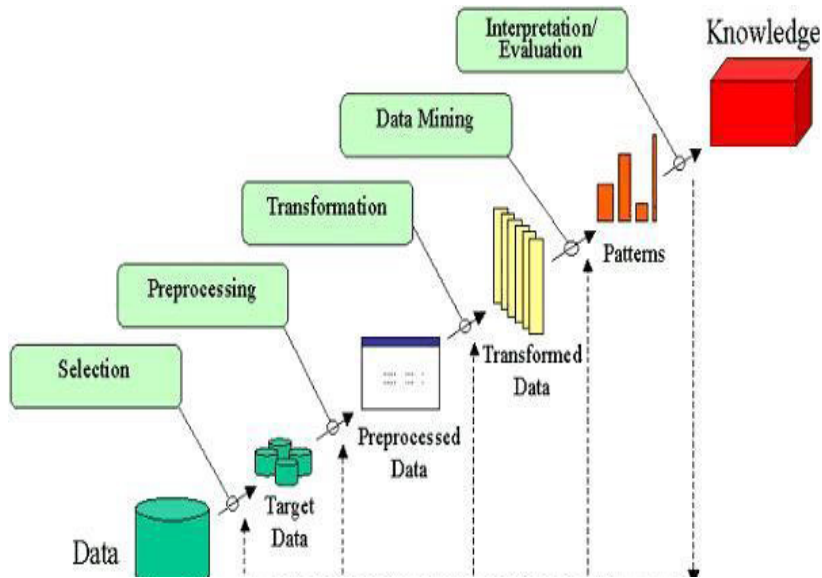


Figure2. Iterative Process in KDD

According, to the Figure2 the following are the iterative steps in Knowledge Discovery [3]:

Data selection (where data relevant to the analysis task are retrieved from the database)

Data cleaning (to remove noise and inconsistent data)

Data integration (where multiple data sources may be combined)

Data transformation (where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operations)

Data mining (an essential process where intelligent methods are applied to extract data patterns)

Pattern evaluation (to identify the truly interesting patterns representing knowledge based on interesting measures)

Knowledge presentation (where visualization and knowledge representation techniques are used to present mined knowledge to users)

5. DATA MINING FUNCTIONALITIES

In general data mining functionalities used to specify kinds of patterns to be found in data mining tasks [3]. Such tasks are classified into two categories:

5.1. Descriptive:

Descriptive mining tasks characterize properties of the data set. Clustering, summarization, association rules and sequence discovery are usually viewed as descriptive in nature.

5.2. Predictive:

Predictive mining tasks perform induction on the current data to make predictions. Predictive modeling may be made on the use of other historical data. Classification, Regression, Time series analysis and prediction are predictive in nature. The Figure-3 shows task in data mining:

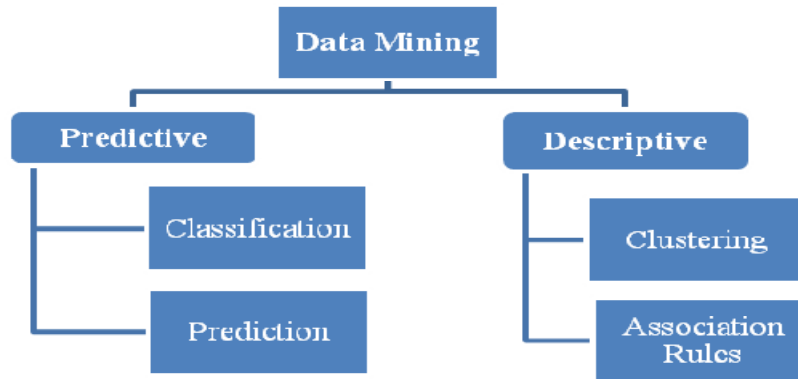


Figure 3. Task in Data Mining

5.2.1. In data mining the data is mined using two learning approaches [6]. They are

A. Supervised learning:

Supervised learning is also called directed data mining. The variables under investigation are split in two groups. 1. Explanatory variables, 2. one or more dependent variables. The goal of analysis is to specify a relationship between the dependent variables and explanatory variables as we do it in regression analysis. To proceed with directed data mining techniques, the value of the dependent variable must be known for a sufficiently large part of dataset.

B. Unsupervised learning:

In unsupervised learning, all the variables are tested in the same way there is no distinction between dependent and explanatory variables. However, in contrast to the name undirected data mining, still there is some target to achieve. This target might be as data reduction as general or more specific like clustering. The divided line between unsupervised learning and supervised learning, in the same way distinguishes discriminate analysis from cluster analysis. Supervised learning requires target variable which should be well defined and sufficient number of its value are given. In unsupervised learning target variable, has only been recorded for the small number of cases as the target variable is unknown.

5.3. Tasks in data mining:

Data mining was divided for the specific classes of six activities or tasks as follows. In this classification, estimation, prediction belongs to supervised learning and association rule, clustering, description & visualization belongs to unsupervised learning.

5.3.1. Classification: classification consists of examining the features of a newly presented objectset consist of reclassified examples. The tasks build a model that can be applied to unclassified data which classify it.

5.3.2. Estimation: Estimation deals with continuous value outcomes. Given some input data we use estimation to come up with a value for some unknown continuous variables.

5.3.3. Prediction: Any prediction can be thought of as classification or estimation [7]. Predictive tasks feel different because the records are classified according to some predicted future behavior or estimated future value. With prediction, the only way to check the accuracy of the classification is to wait and see. The historical data is used to build a model that explains the current observed behavior. When this model is applied to current inputs, the result is prediction of future behavior.

5.3.4. Association: An association rule is a rule which implies certain association relationships among a set of objects (such as “occur together”(or) “one implies the other”) in a database. Given a set of transactions, where each transaction is a set of literals (called items), an association rule is an expression of the form $x \rightarrow y$, where x and y are set of items. The intuitive meaning of such a rule is that transactions of the database which contains x tends to contain y .

5.3.5. Clustering: Clustering analysis can be used as a standalone data mining tool to gain insight in data distribution, or a preprocessing step for the other data mining algorithms operating on the detected clusters. Many clustering algorithms have been developed and are categorized from several aspects such as partitioning methods, hierarchical methods and grid based methods. Further data set can be numeric or categorical. Clustering is the task of segmenting a diverse group into number of similar subgroup (or) clusters. The difference between clustering & classification is that cluster does not rely on predefined classes. In clustering, there is no predefined class, the records are grouped together based on self-similarity.

5.3.6. Description & Visualization: Data visualization is a powerful form of descriptive data mining. It is not always easy to come up with meaningful visualizations, but the right picture really can be worth a thousand association rules, since the human beings are extremely practiced at extracting meaning from visual scenes.

Knowledge discovery goals are defined by the intended use of the system. There are two types of goals. They are verification and discovery. With verification, the system is limited to verifying the user's hypothesis. With discovery goal is further divided into prediction, where the system finds patterns for predicting the future behavior of some entities and description, where the system finds patterns for presentation to user in human understandable form.

6. REVIEW

Steel consists of alloy iron, carbon and manganese with small amount of silicon, phosphorous and Sulphur. Steel production stages: Heating, cooling, melting, solidification. India is 3rd largest of crude steel (up from eighth in 2003) and is expected to become 2nd largest producer in near future. Most industrial applications of data mining in steel industries is system modeling, approaching new manufacturing technologies and to improve the quality of products, anti-corrosive properties, the galvanized steel is a product experiencing an increasing demand in multiple sector. Nonlinear modeling comprises important techniques that have reached broad applicability due to efficiency in speed. The aim of data mining analysis was to predict the yield strength, elongation as functions of large number of variables, including the chemical composition, heat treatment, strip speed in

the annealing process. It is also being used in industrial process optimization and control. This extensive study is to analyse existing techniques used in blast furnace in steel industries.

The main objective of a blast furnace is to reduce chemically and convert physically iron oxides into liquid iron, which is called hot metal. Hot metal manufacturing process consumes about 70% of the entire energy of steel manufacturing integrated route. Besides, social and industrial needs for iron steel, high prices of raw materials and reducing agents have also increased the necessity to model this complex process to increase productivity and reduce cost [2]. To implement this, data mining techniques have been tried in many stages of blast furnace operations.

According to G. J. Zheng, W. Zhang, P. Hu & D. Y. Shi, [8] in this process of hot forming, many design variables have effect on the result in a different and complex way such as geometry features and forming process parameters. It is difficult to understand the relationship between design variables and results, which is very important to guide the design. In this paper data mining was introduced to explore the influence of the past geometric feature and the hot forming parameters on hot forming results of an automobile B-pillar model and the optimum parameter ranges were determined. First several variable parameters were selected and 100 groups of experimental data were generated with Super Latin Method and then finite element method analysis results were calculated respectively. Next analysis and evolution of simulation were carried by making use of Decision Trees (DT) algorithms. Finally, a series of B-pillar hot forming rules were refined, such as initial temperature of the sheet metal should be controlled between 720°C to 800°C. Therefore, a real B-pillar model was designed to test the rules and result was correct and effective.

In this work [9] authors developed a decision support system that can determine corrosion and project the time for corrosion growth for maintenance using probabilistic modelling approach. In this work, Data Mining based corrosion control process is introduced, the process is aimed at preventive and predictive maintenance instead of the existing practice which uses fixed scheduled inspection and using corrosion sensors continuously to facilitate preemptive actions. The use of "Data Mining" process includes analyzing historical data collected over time about known metal thickness previously exposed to environment factors such as rainfall, temperature, humidity & so forth. The results of the analysis provide support for decision making by taking proactive and knowledge driven decisions on corrosion control. Classification method like Bayes theorem approach, Neural networks were used to model uncertainties in corrosion occurrence considering both knowledge uncertainties and data uncertainties to make information decisions. The probabilistic modeling is used to develop a data mining method that can be applied to obtain information from a database on metals previously exposed to an atmosphere environment.

The authors Mohamad Saraee, Mehdi Moghimi & Ayoub Bagheri suggests [10], the annealing process is one of the important operations in production of cold rolled steel sheets, which influences the final product quality of cold rolling mills. In their process, cold rolling coils are heated slowly to a desired temperature and then cooled. Modelling of annealing process (prediction of heating and cooling time and trend prediction of coil core temperature) is very sophisticated and expensive work. Modeling of annealing process can be done by using thermal models. In this paper modelling of annealing process is proposed by data mining techniques due to high speed in data processing, acceptable results is obtained and its simplicity to use it. In this study, they proposed a method for modeling the annealing process by using data mining techniques. After testing different techniques of data mining, the feed forward back propagation neural network is selected to predict heating and cooling times and temperature of coil core during annealing process. The result of predictors applied neural networks for modeling annealing process were accurate enough, but while we use a larger data set accuracy of predictors will be improved. The present method is applicable to predict the behavior of processes that cannot be described by any analytical or physical equations. The other techniques such as

classification, different regressions or clustering method can be applied to optimize the input parameters of annealing process for maximizing productivity of annealing operations.

According, to Sayed Mehran sharafi & Hamid Reza Esmaeily[11], applying data mining methods to predict defects on steel surface was challenging one. In the steel industry, especially alloy steel, creating different defected product can impose a high cost for steel producers. One common defect in producing low carbon steel grades is Pits & blisters defect. Its drawback is waste of time and cost to eliminate this drawback, we need to grind the surface of the product. In some cases, the severity of defects may lead to scrap part of the product. Grinding cause waste of time and cost of production will be increased. Incidence of defects is related to several factors including material analysis and production processes. In this study authors created a model to predict this fault with data mining methods including decision tree, neural network respectively. They applied these techniques to the data collected from Iran alloy steel company. The model created using decision tree has higher accuracy. The benefits of this model are

* Reduces the process time and energy cost by omitting grinding step.

* There is an optimal tool for predicting defect and applying for producing products with no defects.

According, to the authors [12] this paper revealed a data mining approach for variable selection and knowledge extraction from the dataset. The approach is based on unguided symbolic regression (every variable present in the dataset is treated as the target variable in multiple regression runs) and a novel variable relevance metric for genetic programming. The relevance of each input variable is calculated and a model approximating the target variable is created. The genetic programming configurations with different target variables are executed multiple times to reduce stochastic effects and the aggregated results are displayed as a variable interaction network. This interaction network highlights important system components and implicit relations between variables. The whole approach is tested on a blast furnace dataset, because of the complexity of the blast furnace and the many interrelations between the variables.

Many variables in the blast furnace process are implicitly related, either because of the underlying physical relations or because of the external control of blast furnace parameters. Examples for variables with implicit relation to other variables are the flame temperature or hot blast parameters. Usually such implicit relation was not known a-priori in data-based modeling scenarios but could be extracted from the variable relevance information collected from multiple GP runs.

Using an unguided symbolic regression data mining approach several models have been identified that approximate the observed values in the blast furnace process rather accurately. The experiments also lead to many number of model describing several components of the blast furnace. The generated model is used to extract information about implicit relations in the dataset to further reduce and disambiguate the set of relevant input variables.

The next paper also describes application of symbolic regression on blast furnace using temper mill dataset[13]. It reveals the application of an adapted symbolic regression system on two different datasets, the first one contains measurements from blast furnace process which is most common to produce hot metal(liquid iron). Although chemical and physical process are well understood the heat loss in certain areas of the furnace are not completely understood. The knowledge about such relationships can be used to optimize the blast furnace process and therefore modeling the blast furnace process based on collected real world data is of specific interest. Regression analysis is a sub field of data mining attempting to reveal knowledge contained in given set. In this work, symbolic regression is performed using a tree-based on genetic programming system to evolve mathematical formulas. Genetic programming [14] is an

evolutionary algorithm that produces programs to solve a given problem. In this modeling approach three algorithmic aspects are incorporated and compared to a standard symbolic regression approach. Precisely the authors tested the effects of off-spring selection, using the coefficient of correlation R^2 as fitness function and additionally sample the evaluated samples or fitness cases of individual. Off-spring selection is an additional selection setup in genetic algorithms and genetic programming that is applied after recombination and mutation. The fitness of an individual in symbolic regression analysis is commonly calculated as a Mean Square Error (MSE) between the predicted value of the model and the observed value of the target variable. The genetic algorithm is faster than the genetic algorithm without off spring selection. In this contribution three algorithmic adaptations, off spring selection, coefficient of determination R^2 as fitness function and sampling, to a rather standard symbolic regression system have been investigated. The effects of combining these adaptations have been demonstrated on real world data sets from two steel production processes. The best improvements in terms of quality were achieved due to the use of R^2 as fitness function.

According, to Jong-Hag Jeon [15], this presentation is an example of Data Mining Applications at steel making factory. Engineers who are responsible for quality control encounter two kinds of difficulties in solving problems using statistical methods. Hot blast stoves are important to energy consumers. Hot blast sensible heat supplies 10%-15% of the total energy consumed by the blast furnaces. Reduction in energy consumption for hot blast heating has important impact on hot metal production cost.

The productions of industrial blast furnaces require the constant delivery of hot blast air from blast furnace stoves. The blast air must be supplied with temperature and flow-rate as close as set point. In the long term, the control must also meet the conflicting objectives of thermal efficiency, cyclic stability and load response speed. Optimizing combustion conditions contributes largely not only to the prolongation of hot stove service life, but also increases in thermal efficiency of the stove during operation. This paper shows the hot stove combustion analysis both theoretically and experimentally to optimize combustion condition preventing excessive development of combustion flames as well as the abnormal heating effect mainly after burning. In addition, the development feed forward guidance simulator is based on Neural Network Models and multi-stage combustion pattern control systems developed through Data Mining. The introduction of these systems required only minimal operation costs, decreasing the hot blast stove energy cost approximately 15%. The author had developed this system with SAS for no.3 blast Furnace in Pohang steel works, Korea.

High productivity is achieved at Pohang No.3 blast furnace through the following process:

- *Optimization of heat efficiency: automatic combustion control system, calorie control system.
- *Application of new control concept: Multi stage combustion control system.
- *Optimization of operation condition using data mining.
- *Increase of model accuracy using Neural Network Model

POSCO blast furnaces will focus on high productivity operation including high efficiency and long campaign life through consistent technology development.

The following paper revealed [16] the use of data analytics tools for predicting the fatigue strength of steels. Several physics-based as well as data-driven approaches have been used to arrive at correlations between various properties of alloys and their compositions and manufacturing process parameters. Data-driven approaches are of significant interest to material engineers amount to Neural Networks, Reduced error pruning trees, MS model trees (Reconstruction of Quinlan's algorithm). In this study, a range of advanced data analytics techniques, typically involving a combination of feature selection and regression methods, have

been successfully employed and critically evaluated for the problem of fatigue prediction of different grades of steels.

They tested with 12 predictive modeling techniques in this study, which includes the following: Linear regression, pace regression, regression post non-linear transformation of select input variables, Robust fit regression, multivariate polynomial regression, K-nearest neighbor (KNN) modeling, Decision table, Support Vector Machines, Artificial neural networks, Reduced error pruning trees. Thus in particular, neural networks, decision trees and multi-variate polynomial regression were found to achieve a high R^2 value of greater than 0.97, which is significantly better than what has been previously reported in the literature. It is very encouraging to see that despite the limited amount of data available in this dataset, the data-driven analytics models could achieve a reasonably high degree of accuracy.

The authors suggest that [17] Iron and steel melting is an extremely complex physical and chemical process. Blast furnace is main method of iron and steel melting .To obtain the knowledge has become more and more difficult task on manual analysis of the data. Matured theory like data mining and knowledge discovery are applied to acquire new knowledge. Data mining methods are used in the blast furnace production control. Firs the authors introduced clustering algorithms. Then clustering analysis of blast furnace operation parameters was carried out by K-means clustering. Analysis and comparison of practical data was conducted to determine the optimal cluster number of this algorithm for blast furnace parameters analysis, and this yielded the ideal operating value for the parameters. The optimal threshold for blast furnace parameters were determined through statistical analysis, repeated experiments and field assessment, and the difference between blast furnace state as estimated and the practical one analyzed. Finally, the factor analysis method to reduce the dimension of parameters was successful and mining test of Tangshan iron and steel shows that the method is effective in practical application.

According, to the authors [18] the blast furnace for steel making process is examined. As the process is conducted at extreme conditions, it is impossible to observe what is occurring inside the blast furnace. Nevertheless, optimizing the process would greatly improve the overall process and quality of the final product. By implementing the ANFIS (Artificial Neural Fuzzy Inference System) model on collected blast furnace data, they aimed at optimizing the blast furnace process more specially predicting the performance indicator nc_0 , which describes gas utilization rate in the furnace. These studies used linear modeling techniques to study the effects of several explanatory variables of the silicon content in the hot metal part of the output. The results were promising but were outmaneuvered by nonlinear soft computing techniques. The nonlinear soft computing techniques have been applied to blast furnace process modeling. These studies have used neural network-based hybrids to model furnace performance indicators with good results. A well-known data mining framework, CRISP-DM, was used to guide the analysis and modeling process. The ANFIS model was completely data- driven except for some expert knowledge that was utilized in the pre-processing stages and in the construction of final input series, were several series consists of mathematical combinations of two or more raw input series. The adequate performance of ANFIS combined with a proposed pre-processing approach resulted in a system which is feasible in real world applications.

According, to the authors this study [19] described one of possible approaches to deoxidant cost optimization in steel production based on the expert system. Education of the system is based on successfully completed heats. Decision support Systems(DSS) that uses intelligent conclusion (including fuzzy)are used in industry, particularly in the ferroalloys production. The authors used Naïve Bayes network and decision tree(ID3). But decision tree did not go well as naïve Bayes. So, authors used Naïve Bayes network for proposed production model due to its speed, simplicity and ease of interpretation of results. However, this method does not allow the direct processing

of continuous variables – they must be divided into number of intervals to discretize values. DSS on Bayesian network shows higher accuracy of approximation at the training and test sample. According, to the authors [20] the products manufactured by manufacturing company products may vary in quality than expected quality in terms of many parameters. Due to rapid development in data mining and machine learning industries started to do analysis in data mining. In steel quality of product starts from blast furnace process, so they started using data mining techniques to improve the product quality. This existing work used random forest method to detect mild steel defect diagnosis. The Random forest technique is not used widely in steel industry. This paper taken initiative to do it and they found that random forest technique gives more accuracy (97%) than ANN and single decision tree. This paper focused on enhancing the accuracy and efficiency [21] with which operating set points are calculated for a Continuous Annealing Furnace (CAF) on a Hot Dip Galvanizing Lines(HDGL). Recent papers have proved that data driven models based on artificial Intelligence are a good alternative for developing accurate prediction models in the steel industry. To take advantage of these techniques in the modelling of the galvanizing process, the main requirements are the availability of historical data, detailed knowledge of the chemical composition of steel and time to train the models. Now Ensemble Methods give high overall performance in predicting system set point. Ensemble Methods are built using a set of model and the final output is a combination of the outputs of each individual model.

In this paper these Ensemble Methods Additive Regression(AR), bootstrap aggregating (bagging) are developed to adjust CAF temperature settings in an HDGL. Additionally, five data driven models least median squared linear regression, linear regression, Quinlan’s improved M5 algorithm(M5P), multilayer perceptron neural network, Support Vector Machine(SVM) are chosen from literature as basic components of ensembles. The results for furnace temperature set point predictions obtained with this method highlighted the benefits of using EM rather than other data driven models. Multilayer perceptron with basic learner (MLP with BP) worked well in this model.

6.1 According to thereview [22] the Figure4 and figure 5 reflects the percentage of application of data mining in steel industries :

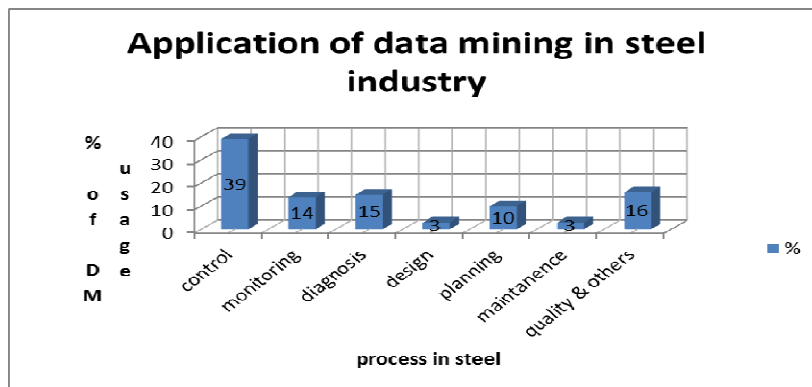


Figure 4. Application of data mining in steel industry

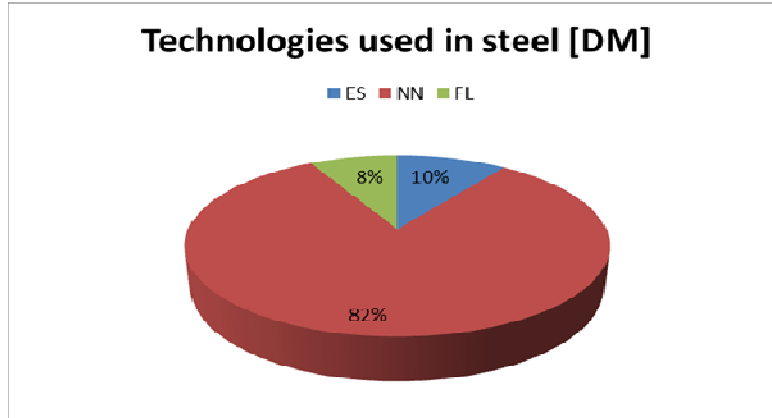


Figure5. Technologies used in Steel [DM]

Abbreviations used in this chart-

DM – Data Mining
ES – Expert System

NN-Neural Network
FL-Fuzzy Logic

7.CONCLUSION

Data mining techniques in steel industries shows excellent results when it is used for the following purposes

- Product quality monitoring
- Process monitoring
- Maintenance strategies

Also, it seems to be clear that data mining can be improved by using numbered modelling simulation where real time data are not available. In all cases, the impact of these technologies in modern industrial process make a requirement for companies to be aware of this to extract knowledge from the process and can improve them. The intension of this survey is to present a review of current work related to usage of data mining techniques in steel industry in Iron making (blastfurnace) and in steel making process. Further research directions will continue in this field. Especially Random Forest and yet more machine learning techniques can be tried, since these techniques are applied in less frequencies than Artificial neural networks, Fuzzy logic techniques in steel industries modelling process.

REFERENCES

- [1] A.K. Choudhury, M.K. Tiwari and J.A Harding, (2009), 'Data Mining in Manufacturing: A review based on the kind of knowledge'. Wolfson school mechanical and manufacturing engineering, Loughborough university, Loughborough, Leicestershire, UK, Journal of Intelligent manufacturing, 20(5), pp. 501-521.
- [2] Rosiane Mary Rezende Faleiro, Claudio Musso Velloso, Luiz Fernando Andrade De Castro, Ronaldo Santos Sampaio, (2013), 'Statistical modeling of charcoal consumption of blast furnace based on historical data: Journal of Materials research and technology', 2(4), 303 – 307.
- [3] Jiawei Han, Micheline Kamber, Jian Pei, (2012), Data Mining: Concepts and Techniques, Third Edition, USA, Morgan Kaufmann Publishers.
- [4] Nine law of Data Mining by Tom Khabaza (<http://www.wkdnuggets.com/2015/16/nine-laws-datamining-part-1.html>).

- [5] Hand D. J, Manila H, & Smyth, (2001): Principles of Data mining, MIT press, Cambridge, Massachusetts. ISBN-262-08290-X
- [6] Aastha Joshi and Rajneet kaur, (2013), A review: 'Comparative study of various clustering techniques in data mining', International Journal of Advanced Research in Computer Science and Software Engineering, Vol 3, 2277-2288.
- [7] Manisha Verma, Mauly Srivastava, Neha Chack, Abul Kumar Diswar, Nidhi Gupta, (2012), 'Comparative study of various clustering algorithms in data mining'. International journal for engineering research and applications, Vol 2, Issue 3, pp. 1379-1384.
- [8] G.J Zheng, W. Zhang, P. Hu & D.Y SHI, (2015), Optimization of hot forming process using DMT and Finite element method, International Journal of Automotive Technology, Vol 16, no.2, pp: 329-337.
- [9] Stephen Dapiap, Gregory Wajiga, Michael Egwurube, Musa Kadzai, Nathaniel Oye & ThankGod Anazodo, (June 2015), Corrosion Control Approach using Data Mining, International Journal of Computer Science & Information Technology (IJCSIT), vol 7, No 3.
- [10] Mahamad sarae school of computing, science and Eng., university of Salford, greater Manchester, UK, Mehdi Moghimi, Dept. of Elec. & computer Eng., Islamic Azad university, Najafabad branch, Isfahan, Iran, Ayoub bagheri, Dept. of Elec and computer Eng, Isfahan university of technology, Isfahan Iran, (2011), Modeling Batch Annealing Process using Data Mining Techniques. ACM journal.
- [11] Sayed Mehran Sharafi, Hamid Reza Esamaely, (2005-2010), Applying data mining methods to predict defects of steel surface, Journal of theoretical and applied information technology, [www.jatit.org].
- [12] Michael Kommenda, Gabriel Kronberger Christoph Feilmayr and Michael Affenzeller, (23 Sep 2013), Data mining using unguided symbolic regression on a blast furnace dataset, arXiv:1309.5931v1 [cs.NE].
- [13] Michael Kommenda, Gabriel Kronberger, Christoph Feilmayr, Leonhard Schickmair, Michael Affenzeller, Stephan Winkler and Stefan Wagner, Application of symbolic regression on blast furnace and temper mill datasets, [n.d].
- [14] John R. Koza, Consulting Associate Professor in computer science department Stanford university, Genetic programming: On the programming of computers by means of natural selection, the MIT press [1992].
- [15] Jong-Hag Jeon, POSCO, Pohang South Korea, Data mining application of six-sigma project, SUGI 29 solutions, paper 186-29.
- [16] Ankit Agarwal, Parjit D Deshpande, Ahmet Cecen, Gautham P Basavarsu, Alok N Choudary and Surya R Kalidindi, (2014), Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters, Integrating materials and manufacturing innovation, 3:8, A springer open journal.
- [17] Fuxing Yu, Yina Suo, Xin Zang, Aidind Yan, Fulong Liu, (2013), Data mining in blast furnace smelting parameter, Applied mechanics and materials, vol. 303-306, pp 1093-1096.
- [18] Bjork, Holopainen, Wikstron, Saxen, Carelsson and Sihdonen, technical report number 1094, (Nov 2013), Analysis of blast furnace time series data with ANFIS: Turku center for computer science [TUCS].
- [19] Zheldak T.A, Slesarev V.V, Volovenko D.O, (2013), Knowledge-based intellectual DSS of steel deoxidation in BOF production process, American Journal of Mining and Metallurgy, Vol. 1, no.1, 7-10.
- [20] Veena Jokhakar, S.V Patel Ph.D., (March 2015), A Review of Business Intelligence Techniques for Mild Steel Defect diagnosis, International Journal of Computer applications (0975 – 8887), volume 113 – No 10.
- [21] Sanz-Garcia, F. Antonanzas-Torres, J. Fernandez-Ceniceros & F.J. Martinez-De-Pison (2014), Overall models based on ensemble methods for predicting continuous annealing furnace temperature settings, Iron and Steel Making, vol. 41, issue no 1.
- [22] Radu Platon & Mouloud Amazouz, From Report CETC – Varennes September 2007 -141 (TR), Application of data mining techniques in Industrial Process Optimization, Prepared by CANMET energy Technology Centre, <http://www.nrcan.gc.ca/2007-141e>.