Automobile Industries using Data Mining and Predictive Analytics: An Industry 4.0 Approach

Rani Kumari
School of Computer Science & Engineering
Galgotias University, Greater Noida, Uttar Pradesh, India

Dr. Kavita Saini
School of Computer Science & Engineering
Galgotias University, Greater Noida, Uttar Pradesh, India

Abstract- The whole world widely use the current trend called Industry 4.0 in manufacturing Industry for smarter decisions, by collecting data smartly, making correct decisions and executing decisions without any doubts, by using the most advanced technologies, the procedures of collecting and interpreting data will be easier. This overall consciousness gives Industry 4.0 the most important aspect of artificial intelligent functions to make the industry even smarter.

The Automobile Industry Framework use different algorithms including Classification, regression and association to take decision. The predictive analytics and maintenance (PdAM) approach help industry to identify pattern by using sensors to indicate changes in equipment condition, typically wear and tear on specific equipment before they get failure.

In this paper a framework PdAM is proposed which has various phases. The phase one is about how to collect data from different resources i.e. sensors, IOT, Data ware house, files and so on. While phase two is Data pre-processing and will be achieve be done by using Data cleaning, Data transformation and Data reduction. Phase three talks about decision of algorithms to be used. After decision of algorithm to be used development of predictive maintenance model takes place in phase four while in phase five Training and Testing of model to generate precise result will take place.

Keywords – Data Mining, Industry 4.0, PdAM, Predictive analytics, Automobile Industry, CRIPS

I. INTRODUCTION

In this era, the first step to make an industry smarter is use Predictive analytics and maintenance (PdAM). The current trend called Industry 4.0 is widely used in manufacturing factories to becoming smarter. Accordingly to prediction industries take decision to increase productivity, operation efficiency and better safety. PdAM collect data like temperature, and sound using special sensors (e.g. ultrasonic or vibration sensors) to identify the patterns and state of the current Machine. For smart manufacturing for Automobile Industries framework is proposed where different algorithms like Classification, regression and association algorithms. In term of maintenance try to make industry an intelligent by collecting a lots of data through sensors and use it to make decision about repairs and upkeep. Predictive maintenance systems are used machine learning to determine when and what maintenance work is to be done.

II. BACKGROUND

Now, this doesn't mean we weren't using data before-We've been gathering data for decades. The difference is simply the sheer volume of data available and the new methods we have for handling it all. IOT(Internet Of Things) with CPS(Cyber physical System) and cloud-based processes allow us to gather and interpret data in ways that weren't possible before, and the impact of those technologies is being felt in every aspect of manufacturing, from

production to maintenance to marketing, and even then on to the final products create. In this research try to focus on predictive maintenance by using machine learning.

III. Research Aim and Objectives of Predictive Analytics

- Early awareness of the health of the machine
- Intervention before the machine is damaged
- Reduce downtime
- Improve safety
- Understand the root causes of failure
- Increase productivity and operational effectiveness
- Increase revenue

IV. The major Inputs on which basis PdAM take decision of maintenance

- Temperature
- Sound
- Vibration
- Pressure
- RPM
- Machine details
- Productivity details etc.

V. Methodology Frame work

This research follow the methodology is Cross Industry Standard Process for Data mining (CRISP-DM). CRISP-DM methodology provides a structured approach to planning a data mining project. Because one of the most reliable, robust and user-friendly is the CRISP-DM technique. It consist six phases.

1. Business understanding
2. Data understanding
3. Data preparation
4. Modelling
5. Evaluation
6. Deployment

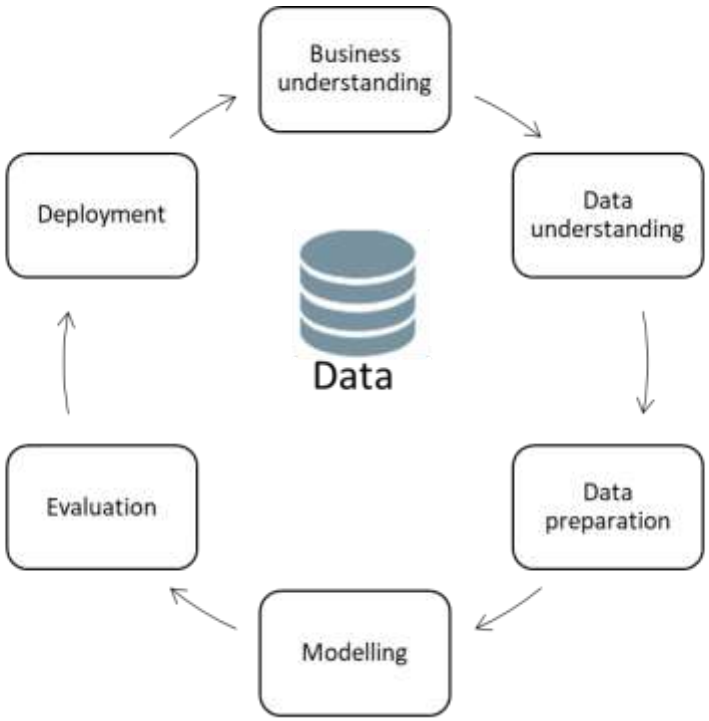

Figure 1. CRISP-DM Process

VI. CRISP-DM Process Sequence and Task

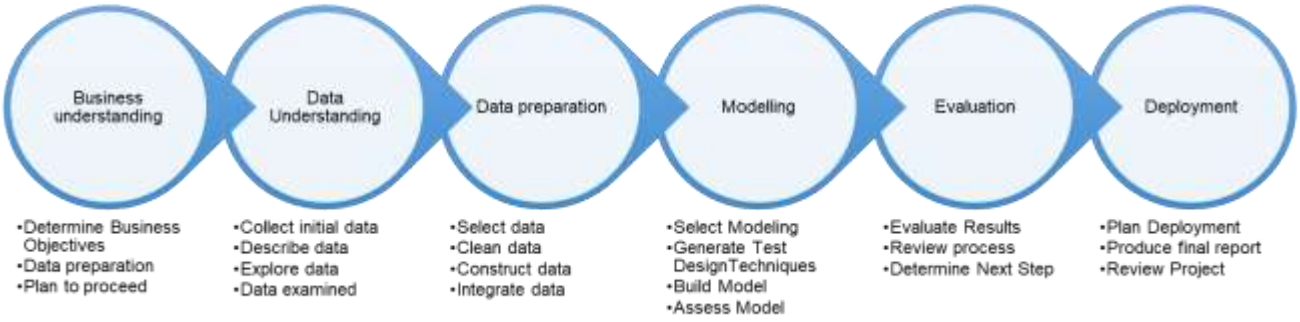

Figure 2. CRISP-DM Process Task

1 Business Understanding

This project starts with the project's goal that is early awareness of the health of the machine, intervention before the machine is damaged, reduce downtime and Improve safety etc included in the first phase “Business Understanding”. Hence, It focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a problem definition and a preliminary plan.

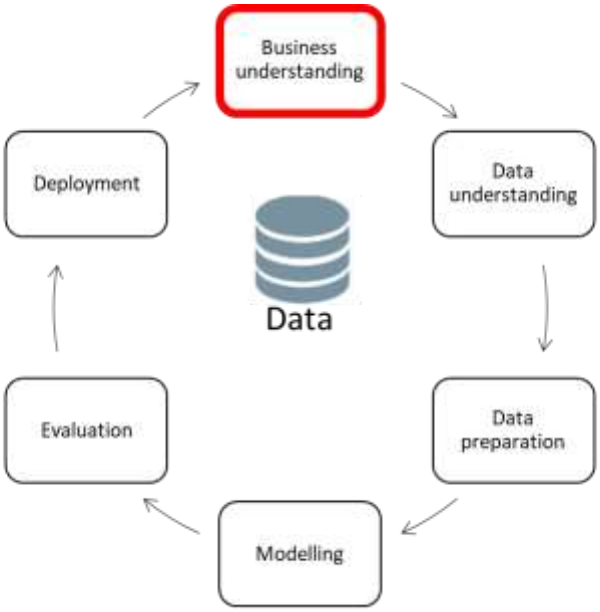

Figure 3. CRISP-DM Process (Business Understanding)

2 Data understanding

Starts with an initial data collection and proceeds with activities in order to get familiar with the attributes. Understand the meaning of each attribute and attribute value in business terms. Hence, acquire the data, explore the data (query & visualization) and verify the quality of data.

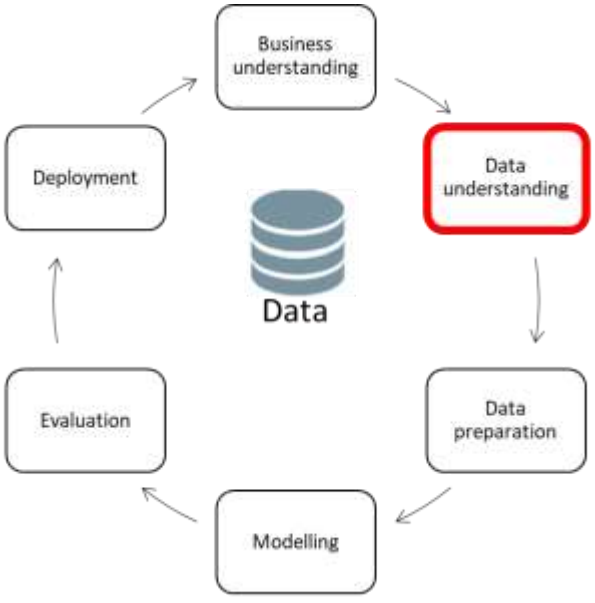

Figure 4. CRISP-DM Process (Data Understanding)

3 Data preparation

The data preparation phase covers all activities that is proper selection, cleansing, constructing and formatting is done to construct the final dataset from the initial raw data. Usually it takes over 90% of the time. Data preparation tasks are likely to be performed multiple times and not in any prescribed order.

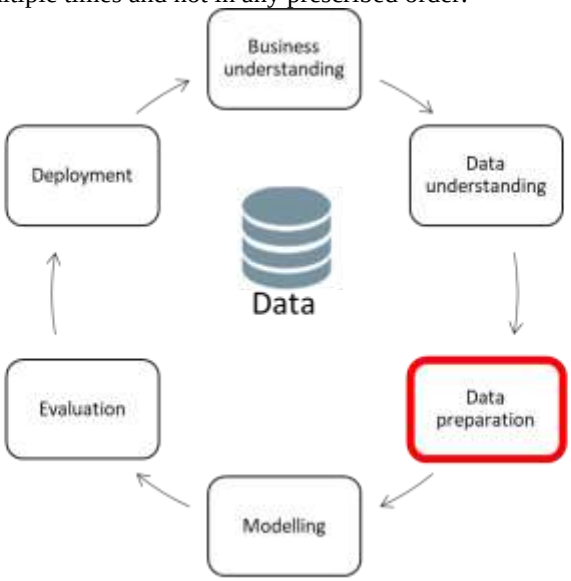

Figure 5. CRISP-DM Process (Data preparation)

4 Modelling

Modelling techniques are selected and applied. Since some techniques like Artificial neural nets, k-nearest neighbors and decision tree have specific requirements regarding the form of the data, there can be a loop back here to data preparation.

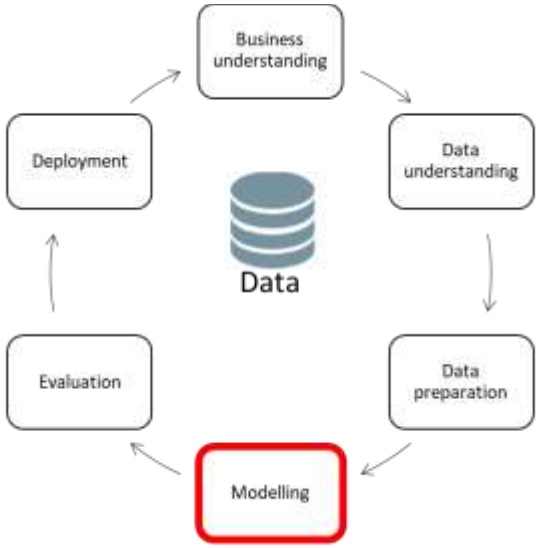

Figure 6. CRISP-DM Process (Modelling)

5 Evaluation

Evaluation tested to ensure they (predictive model) generalize against unseen data and that all key business issues have been sufficiently considered. A key objective is to determine if there is some important business issue that has not been sufficiently considered.

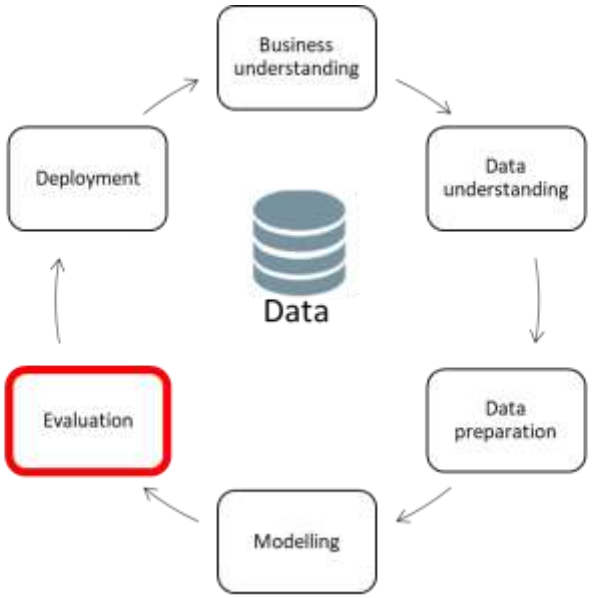

Figure 7. CRISP-DM Process (Evaluation)

6 Deployment

To create a mechanism for the use of that new information in the solution of the original business problem.

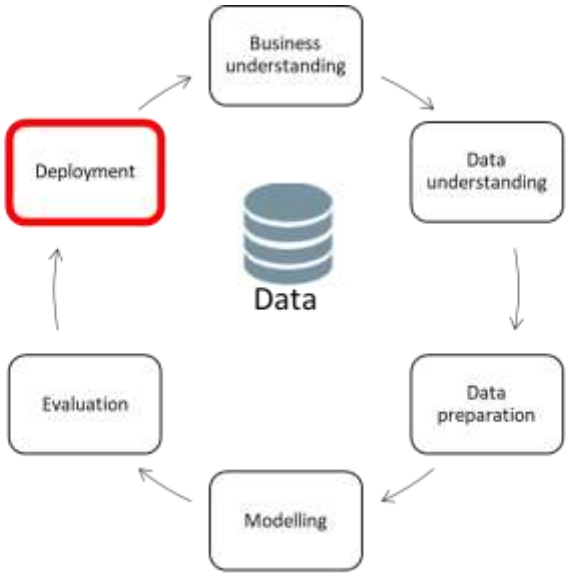

Figure 8. CRISP-DM Process (Deployment)

VII. Data Mining & Predictive Analytics Workflow

Predictive analytics is the process of data analytics to make predictions based on data. This process uses data along with analysis, statistics, and machine learning techniques to create a predictive model for forecasting future prediction. Predictive analytics is incomplete without its immediate predecessor Data Mining. Data mining is the process of discovery useful patterns and trends from the large data sets. With the help of data mining we can retrieve the valuable information from the huge amount of data and make the data usable for analytical purpose. This image show the workflow of model creation and automation.

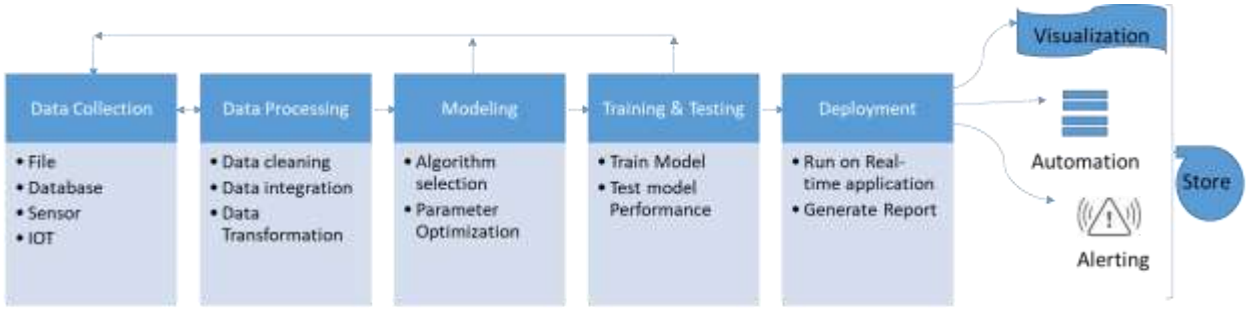

Figure 9. Data Mining & Predictive Analytics Workflow

First offal understand the business objectives, what need to achieve. After that move in the next phase to collect and understand the data from different resources like files, Database, IOT, sensors. There are different type of sensors to collect data. Among them some sensors are

- ***Temperature Sensor***- It measures the changes in the temperature i.e. senses the temperature. Temperature Sensors are used everywhere like computers, mobile phones, automobiles, air conditioning systems, industries etc.
- ***Proximity Sensors***- That detects the presence of an object. Some of the applications of Proximity Sensors are Mobile Phones, Cars (Parking Sensors), industries (object alignment) etc.
- ***Vibration Sensors***- Vibration sensors are piezoelectric accelerometers that sense vibration. They are used for measuring fluctuating accelerations or speeds or for normal vibration measurement.
- ***Accelerometer*** - An accelerometer is an electromechanical device that will measure acceleration forces. The motion sensors in accelerometers can even be used to detect earthquakes.

Other Common datasets and schema

- IOT data (temperature, vibration, sound, voltage, image, etc.)
- Metadata about equipment (e.g. make, model, or revision)
- Usage history(A car driven 100,000 miles is more wear and tear than a car driven 20,000 miles)
- Maintenance data(If an equipment battery is recently changed it has less chance to breakdown compared to other equipment that have not been inspected for a long time)

By using these sensors and other schemes collected data are like:-

- Temperature
- Sound
- Vibration Pressure
- RPM
- Machine details
- Productivity details etc.

Once the data collected the next phase is **data Pre-processing**. Data Preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis. These techniques are used for make data useful-

- ***Data cleaning-*** It focus on handling missing data, noisy data, detection, and removal of outliers, minimizing duplication and computed biases within the data.
- ***Data integration***- when data is gathered from various data sources and data are combined to form consistent data. This consistent data after data cleaning is used for analysis. some of them are
- ***Normalization***- In this process, numerical data is converted into the specified range, i.e., between 0 and one so that scaling of data can be performed.
- ***Aggregation***- this method is used to combine the features into one. For example, combining two categories can be used to form a new group.
- ***Generalization***- In this process, lower level attributes are converted to a higher standard.
- ***Data transformation-*** This step is used to convert the raw data into a specified format according to the need of the model.
- ***Data reduction-*** Redundancy is the nature of data so cannot completely remove the redundancy but here try to minimize the redundancy so that efficiently organize the data and ready to use for analysis and development of the predictive model.

Figure 10. Data Preprocessing

Once the data Preprocessing phase is completed, Start **next phase modeling** i.e. predictive model creation. Here deciding which algorithms and approaches are generally best suited to the specific problem. Because algorithm selection is the most crucial part of the modeling. There are several types of algorithms are available in data mining some of them are listed below-

Table -1 Identification of Data Mining and Predictive algorithms

Techniques	Focus on condition of equipment
Classification	Classify the failure of machine in next '1day', or '1 week' or '1 month'? Accordingly label data i.e Normal instance Failure instance
Clustering	Try to make different cluster of data that get by sensor (e.g. ultrasonic or vibration sensors) like –heat, vibration, RPM etc.
Regression	Determine the remaining life an asset Problem is solved by applying regression technique.
Outlier detection	When data does not fit in any class it is highlighted as outliers.
Sequential pattern	Focus on identifying accurate pattern or abnormal, unusual or irregular patterns recovered by sensor.
Association Rules	In machine one part is related to another part if one is failure than another parts will be failure in near future is identify by applying association rules.

Whenever the data mining and predictive machine learning algorithm is decided, build the Predictive model once it is completed train and test this model using dataset.

Splitting the data-set

The data should be collect from different characteristic of problem so that the algorithm better understand the problem and give good prediction. The decided algorithm model that is going to learn from data to make predictions. Generally split the data-set into 70:30 ratio

- Training data (70%) &
- Testing data (30%)

The 70% of data provided during training phase. It does not matter how much good your algorithm, if less amount of data provide it to learn the pattern then it does not work well. So provide sufficient amount of data to the model.

The 30% of data used during testing phase to check the performance of the model i.e. accurately it work.

However, this Splitting can be varies according to the data-set shape and size.

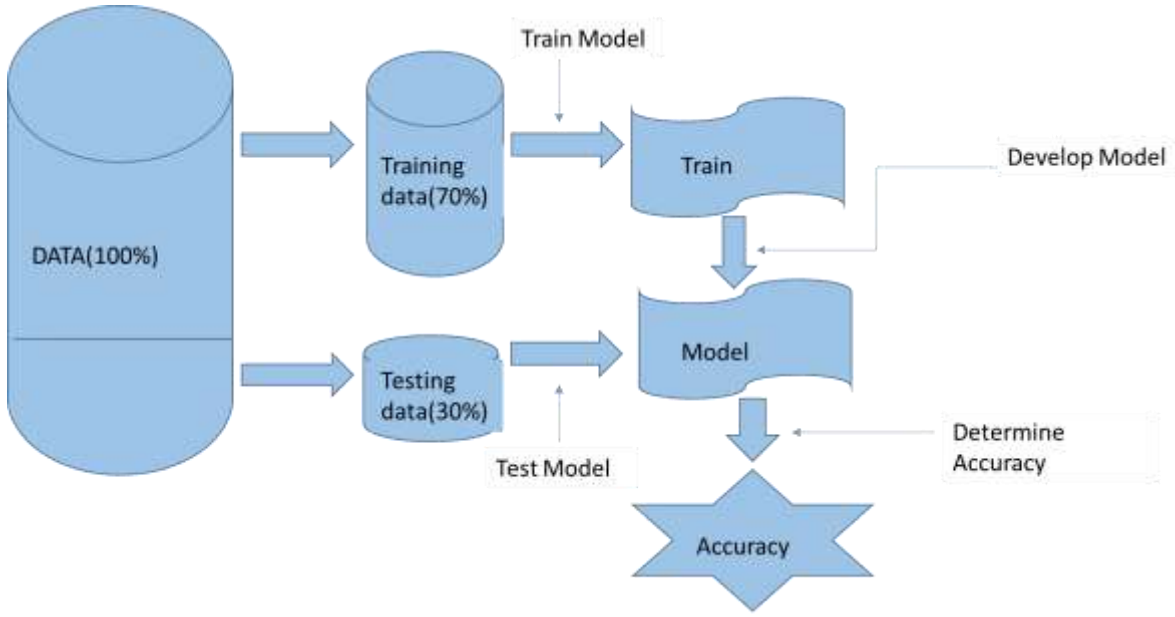

Figure 11. Classification of data set

After determine the accuracy of the model apply the next phase deployment it on the real time application and generate the maintenance report also. The report ensure that a piece of equipment requiring maintenance is only shut down right before imminent failure. The maintenance department takes an action on the generated PdAM repot. This reduces the total time and cost spent maintaining equipment.

Common use cases of PdAM repot are:-

- Predict when an asset may fail in near future.
- Optimize equipment lifetime.
- How to optimize employee productivity.
- How to increase revenue.
- Early warning of anomalies.
- Reduce machine downtime and less wastage of materials.

VIII. CONCLUSION

The arrival of Industry 4.O has provided a new opportunity for PdAM. The Data Mining and PdAM use different techniques to improve productivity and safety are Classification, Regression and Association etc. The PdAM collect data (like-heat, vibrations and RPM etc) using sensor and analyze these data such as vibration analysis, oil analysis, thermal imaging, and equipment observation. During creation of model it is very important phase of selecting algorithm. That is which algorithm is well suitable for these type of data and problem to identify pattern. Once Algorithm is decided then build machine learning model and prepared maintenance report .The maintenance report has a lot of suggestion regarding maintenance (Condition based maintenance and interval based maintenance).the maintenance department work on that suggestions. The PdAM not only provide cost saving but also enables new business models.

REFERENCES

[1] Mr. Sudhir M. Gorade, Prof. Ankit Deo, and Prof. Preetesh Purohit, “A Study of Some Data Mining Classification Techniques”, International Research Journal of Engineering and Technology (IRJET), Volume 04, Issue 04, Apr -2017

[2] Shivangi Bhardwaj,” Data Mining Clustering Techniques – A Review”, International Journal of Computer Science and Mobile Computing, Vol. 6, Issue. 5, pg.183 – 186, May 2017

[3] Amanpreet Kaur,” Predicting Instructors Performance using Data Mining Techniques”, IJESC, Volume 7, Issue No.6,2017

[4] Lakshay Swani1, Prakita Tyagi2,” Predictive Modelling Analytics through Data Mining”, International Research Journal of Engineering and Technology (IRJET), Volume 04, Issue 09, Sep -2017

[5] Steffen Hubera *,” DMME: Data mining methodology for engineering applications –a holistic extension to the CRISP-DM mode”, 12th CIRP Conference on Intelligent Computation in Manufacturing Engineering, July 2018

[6] Prashant Dhingra,” A strategy for implementing industrial predictive maintenance”, October 5, 2018, [Online]. Available: https://cloud.google.com/blog/products/data-analytics/strategy-implementing-industrial-predictive-maintenance-part-i

[7] GRAHAM IMMERMAN, "The Impact of Predictive Maintenance on Manufacturing", July 24, 2018, [Online]. Available: https://www.machinemetrics.com/blog/the-impact-of-predictive-maintenance-on-manufacturing

[8] MARCEL KÖHLER, "Industry 4.0: Predictive maintenance use cases in detail", [Online]. Available:https://blog.bosch-si.com/industry40/industry-4-0-predictive-maintenance-use-cases-in-detail/

[9] Daniel T. Larose, ”Data Mining and Predictive Analytics (Wiley Series on Methods and Applications in Data Mining)” Kindle Edition

[10] Dr. Kavita, “Recent Advances and Future Research Directions in Edge Cloud Framework”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-2, December, 2019,DOI: 10.35940/ijeat.B3090.129219

[11] Dr. Kavita “Evolution of Blockchain Technology in Business Applications” , Journal of Emerging Technologies and Innovative Research (JETIR), Vol.6, Isue 9, Page 240-244, JETIR May 2019, (ISSN-2349-5162)

[12] Jiawei Han, "Data Mining: Concepts and Techniques", August 2000