

Condition Monitoring and Predictive Maintenance of Process Equipments

Manthan Deshmukh^{1,*}, Rohan Dumbre^{1,**}, Shubham Anekar^{1,***}, Heramb Kulkarni^{1,****}, and Sushant Pawar^{1,†}

¹Department of Instrumentation Engineering, Ramrao Adik Institute of Technology, Navi Mumbai

Abstract. Industry 4.0 the proclaimed fourth industrial revolution is unfolding at the moment. It is characterized by interconnectedness and vast amounts of available information. Industrial production has evolved enormously over the last centuries due to modern instruments. Hence issue of the instrument failure is very paramount in any industry. Even if one machine fails it halts the whole production. Overall, it may cost us with more man-hours, project delay, process latency and all this sums up as a huge loss. The life of the instruments should be taken care by continuously monitoring its health. Any faulty or unnatural disturbance in usage of the instrument may lead to its failure. Every instrument needs proper maintenance, even with the slight negligence towards the anomaly it may lead to instrument failure. In, predictive maintenance historic data is utilized and analyzed with the help of advance analytics and modelling techniques using Machine learning, moreover we can predict failures and can schedule the maintenance beforehand and predict failure in advance. With the help of relevant sensor dataset, we can estimate the remaining runtime of the instruments. This maintenance approach helps to lower the costs which are incurred due to system shut downs. It also ease the scheduling and maintenance activities. In this work, three different industrial case studies are considered like shell and tube type heat exchanger, plate type heat exchanger, and semiconductor manufacturing process. Here the predictive maintenance is carried out for heat exchanger by utilizing the concept of multi linear regression and time series analysis. For the semiconductor manufacturing dataset, support vector machine algorithm is implemented to find out the good and bad quality of semiconductor production slots.

1 Introduction

In the environment of Industry 4.0, proper cost allocation and time management is the primary concern to lead in this competitive Industry [10]. Equipment failure or process shutdown will cost an industry a heavy fortune. To avoid this, proper timely maintenance is especially important. In early days before the digitalization preventive maintenance technique was used, were-in equipment was kept running in the process until failure. This was not a cost saving and time efficient method [11]. Predicting failure well in advance with the help of advance analytics can significantly outperform traditional maintenance approaches. In contrast to preventive maintenance, predictive maintenance can be especially useful as it increases productivity, reduces breakdown and lowers maintenance cost. In process industries, a lot of sensor data is recorded, and it plays an important role in predictive analysis. Using this recorded data and machine learning algorithms like Linear/Multi Linear Regression [1], Logistic Regression [2], Time-Series Forecasting [3], Cluster Analysis, etc we can build a model which will help us with monitoring the health of the equipment and forecasting the future results well in advance and apparently helping us to schedule a

maintenance beforehand. These data have a lot of hidden information which can be an asset to the company. Analysing these data can help to know the behaviour of the equipment. Also, various conditions can be analysed in order to distinguish between operating and abnormal conditions. The data can also be helpful for analysing the process behaviour and can be used to make the process more efficient. Hence data can be the driver for the betterment for the industry and for the future growth.

1.1 Motivation and Background

Everyday a lot of data is generated in every field which varies from business to business. In case of instruments a lot of sensor data is generated through various field instruments. This data can be made to work using predictive modelling techniques [16]. Though traditional techniques i.e., preventive and reactive maintenance are widely used in many industries. A lot of advancements are made, and continuous research is going in improving the quality of data for optimum results. Hence it marks a new way in maintenance approach, better manage inventory, eliminate unplanned downtime and maximize equipment lifetime.

In 1990's Predictive Maintenance came into the picture, still most of the industries, were unaware due to lack of digitization. But now-a-days predictive maintenance is emerging, and various industries are shifting towards it. Predictive maintenance has evolved from the first method that is visual inspection to digitized methods using the

*e-mail: manthandeshmukh100@gmail.com

**e-mail: dumbre1999@gmail.com

***e-mail: anekarshubham50@gmail.com

****e-mail: Kulkarni.heramb@outlook.com

†e-mail: sushant.pawar@rait.ac.in

statistical and analytical techniques based on trend recognition and machine learning [7]. Predictive maintenance mainly focuses on three aspects: 1) Downtime Reduction – Lesser the downtime due to equipment failures, more the productivity. 2) Cost saving – Equipment failure is prevented, by scheduling the maintenance while gradually leads to low equipment expenditure 3) Energy efficiency – Preventing intermittent and unscheduled downtime resulting in maximizing the productivity leading to proper energy management.

1.2 Related Research

Maintenance is an important activity in various process equipment. It plays a significant role in process industries with respect to production, cost, safety. Hence the maintenance was carried out through various methods since olden days. In early days of production technology breakdown maintenance was used and was reactive in nature [5]. Equipment was kept running until its failure occurred and then replaced this is called as reactive maintenance. As it can lead to serious consequences and is an issue of safety. Due to which it led to time-based maintenance, known as preventive maintenance [8]. In this maintenance of equipment has been schedule before its failure but it may happen that the equipment can work for more period than its maintenance schedule this led to maintenance cost, unnecessary utilization of manpower and equipment life may also get affected. This led to the evolution of predictive maintenance, in which maintenance is schedule well before its failure with the help of historical data and continuous monitoring of equipment to detect and verify the defects [14]. Only when a defect is detected, the maintenance is scheduled. Till the early 90s, spreadsheets were used to obtain approvals for condition-based maintenance programs [13]. Now a days, this is no longer the case. The obvious benefit of predictive maintenance is that it maximizes runtime, reducing maintenance cost, operations, long shutdowns, increases plant safety [18]. Repairs can be carried out just before a breakdown. That represents a major advantage, knowledge shows that unplanned maintenance work led to high downtime cost. Predictive maintenance improves repair time [17]. Since the specific equipment problems are known in advance, maintenance work can be scheduled accordingly. This makes the maintenance work faster and smoother. As necessary action is taken well before failure, there is no secondary damage, thus reduces repair time [12]. As every industry is shifting towards digitalization, it gives an instant boost to productivity helping projects to move faster and manufacturers to hit more aggressive deadlines. So, with this process industries have also marked their way. “Data is the new Oil” and hence if used in a proper way can drive the growth of the industries benefiting cost, safety, life of instruments, sustainability, and future [6]. According to studies it has shown that proper scheduled maintenance reduces costs up to 25 maintenance costs in addition it has increased production in less amount of time [15].

1.3 Contribution and Paper Structure

Predictive analytics is an important concept when it comes to industry 4.0. In order to incorporate in process industries a primary step is the acquisition of large amount of sensor data. Along with data analytics there is also need of various Machine Learning algorithms for prediction [9]. In this paper 1) Predictive maintenance of Heat exchanger is done using Linear/Multi Linear Regression: It is generally used because it gives linear relationship between dependent and independent variables. Fouling factor which is dependent variable in heat exchanger is predicted using independent variables from the dataset 2) Auto Regression Integrated Moving Average (ARIMA) model: This model is used for Time Series Forecasting i.e. to predict the failure for future days based on the historical data. Fouling resistance of heat exchanger for future days was predicted using ARIMA model. 3) Semiconductor industry case study which manufactures equipment wafers, the response of these manufactured wafer was classified using Support Vector Machine (SVM): This algorithm is generally used for regression and classification problems. To classify the targeted class based on the labelled data. In this case the targeted response i.e. the manufactured wafer whether it is “Good” or “Bad” was classified using SVM. 4) Principal Component Analysis (PCA): It is a dimensionality reduction algorithm used to reduce the higher dimensional data into lower without losing the features. The features of the higher dimensional data is reflected in lower dimensions which can be further processed for modelling. In semiconductor case the data was having a higher dimension which was lowered to two dimensions without losing the features and further the targeted response is classified by Support Vector Machine.

Paper Structure is as follows: Section 1 represents the introduction, overall motivation and background of the idea and related research of the same. Actual problem and its brief solution is discussed in section 2. Section 3 describes the various algorithms used to achieve the required results. . Simulations and Results are shown in section 4 which elaborates three case studies and their visualizations. Conclusion and Future Scope talks about the end result and future scope of the predictive maintenance in section 5.

2 Problem Statement

In conventional method of maintenance i.e., Preventive Maintenance unplanned downtimes resulting in the reduction of output, which gradually impacting the direct loss of profit. In, 21st century, plants are increasingly turning towards machine learning (ML) to recognize patterns in sensor data. As a result of shifting towards digitalization, it benefits the business as maintenance activity is predicted well in advance. Hence it reduces the cost of the maintenance. Predictive maintenance is also useful against unnecessary shutdown due to unplanned maintenance. It also increases the safety of the workers because continuous monitoring alerts in advance for any faulty condition and we can stop the workers from coming in exposure to

any hazard. Now-a-days lot of accidents at various process plants are taking place which occurs due to human mistake or due to unplanned maintenance. Traditional methods of monitoring and maintenance did not comment upon idea of lifespan of the equipment, where in contrast Predictive Maintenance can give a predicted picture over remaining useful life and with planned maintenance life of an equipment can be increased with preventing the downtimes. Proper behavioural monitoring of instruments and predicting their remaining useful life in advance can protect the plants from hazards.

- **Monitoring the health of equipment:-** The main aim is to continuously monitor the health of the instruments i.e., whether it is functioning healthy or not. If any fault occurs in the machine or an equipment immediately it is identified by the system and gets notified to the controller in the control room. This can be identified with the help of analysing the historic data and setting the range or the threshold range, where disrupting these ranges will describe the unhealthy situations.
- **To detect anomalies, present in the dataset:-** When the raw sensor data is obtained it contains a lot of anomalies or noise. These should be detected, and data should be cleaned in order to use it for future predictions and visualizations.
- **Predictive analysis of the equipment:-** Every process instrument or machine needs a maintenance for its smooth functioning. But also, if it is carried out anonymously without proper schedule it may lead to decrease the life of machine and lead to maintenance cost. Hence using predictive analysis, we can predict in advance and plan the maintenance well in advance.
- **Comment upon the remaining useful life of equipment:-** The prediction analysis is done using a lot of sensor data, it also includes the timestamp. Hence using the time and the output to predict, we can use time series forecasting algorithm in order to predict the future output i.e., Remaining Useful Life well in advance and hence accordingly predict the maintenance activity.

3 METHODOLOGY

The basic idea of our project is to build predictive maintenance model of process instruments and to predict on the remaining useful life (RUL) using different Machine Learning algorithms. In various process industries various instruments carry out their task efficiently and hence the maintenance of this instruments is especially important for their smooth working and for the protection of assets. As these instruments are a part of different loop processes it has a lot of sensors which are mounted for accurate measurements of different variables. These data are recorded periodically on various timeframes and the same can be used for the predictive maintenance. The data can be processed using different data analytics techniques. The raw data is firstly processed, and noise and pure data is filtered. After we must find the condition indicators to bifurcate between healthy and faulty conditions. The next step is

to build the machine learning model supervised learning algorithms. The model is trained using training dataset, 70 percent of total data is used for training the machine learning model and remaining 30 percent data is used for testing. Based on the training model the testing data or the future predictions can be done. The various types of regression algorithms, classification algorithms are considered in this work such as Linear Regression, Multiple Linear Regression, Support Vector Machine (SVM), Principal Component Analysis (PCA), Auto Regression Integrated Moving Average (ARIMA) and Decision Tree.

- **Linear Regression** are basically of two types simple and among which simple Linear regression is a technique for predetermining the value of one variable dependent on the value of another. The variable to be estimated is known as the dependent variable. It is a basic linear methodology for predetermining a response Y based on a solitary variable X. It has linear or straight relationship among X and Y. So linear Regression is the way toward discovering a line that best fits the data points accessible on the plot, so we can utilize it to anticipate yield esteems for inputs that are absent in the data set we have, with the conviction that those yields would fall on the line. And Multiple regression is a type of linear regression that is expansion of simple linear regression. It is utilized to forecast the value of a dependent term depending on two or more independent terms or variables. Mathematically represented way of the relationship is shown ahead, this regression model is written as $Y = C + m_1X_1 + m_2X_2 + \dots m_nX_n + E$ where 'X_n' is the variable is used to represent the nth predeterminer and 'm_n' compute the dependence between that variable and the response. 'm_n' is interpreted as the mean effect on Y as a single unit increase in 'X_n', holding all other predeterminers fixed.

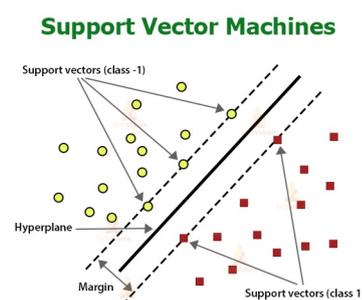


Figure 1. Support Vector Machine representation

- **Support Vector Machine:-** SVM is a sort of machine learning algorithm that can be utilized to solve classification or regression problems as shown in Figure 1. It transforms the data using a technique known as the kernel trick and then finds an ideal boundary between the possible outputs depending on these transformations. It consists of a hyperplane which divides the classes for which classification is to be done[4]. Support vectors are nothing but the data points that are nearer to the hy-

perplane which impact the position and direction of hyperplane. As per the SVM calculation we discover the points nearest to the line from both classes. These focuses are called support vectors. Presently, we process the distance between the support vectors and the line. This distance is known as the margin. We will probably augment the margin. The hyperplane for which the margin is greatest is the ideal.

- **Principal Component Analysis:** - Principal component analysis (PCA) means the method by which its components are computed, and the proper utilisation of these components in understanding the data. It is a dimensionality-reduction process which is often used to shrink the dimensionality of big data sets, by converting a large set of variables into a simple and small one that still contains most of the information in the huge set as shown in Figure 2. PCA is a non-supervised learning approach, as it includes only a set of features X_1, X_2, \dots, X_p , and no other related response Y . PCA is used to produce derived variables for supervised learning approach. It even works as a tool for data visualization. It converts the data in lower dimensions from higher dimensions. Theory here is that every 'n' observation lies in m-dimension, but not every dimension is similarly interesting. Decreasing the quantity of variables of a data set obviously reduces the accuracy, but

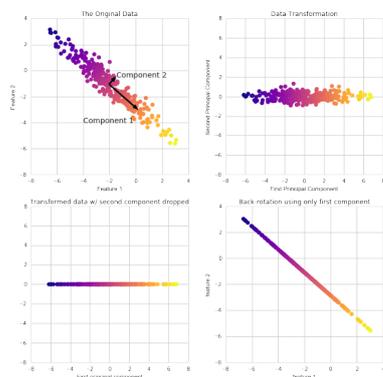


Figure 2. Principal Component Analysis

what can we do here is, in dimensionality reduction we can adjust little accuracy with simplicity. Since small data sets are easy to explore and visualize and do analysing data much easy and fast with machine learning algorithms without external variables to process as shown in Figure 2. PCA asks for a small quantity of dimensions that are interesting like possible, here the theory of interesting is to level by the number that the observations differ along each dimension. Every dimension discovered by PCA is linear combination of the p features.

- **Time Series Forecasting:**-ARIMA stands for Auto Regressive Integrated Moving Average and it is one of the simplest and most powerful machine learning algorithms for forecasting time series. This is the function of combining Auto Regression and Moving Average. ARIMA is a easiest stochastic time series model,

forecasting future time points can be done if use it to train and then forecast. ARIMA can detect complex dependencies as it takes terms and observations of delayed terms. These models depend on regressing a variable on previous values. Auto regression is a time series model that predetermines the value of the upcoming time step by using measurements from past time steps as input to the regression equation. In other words, it predicts t by doing regression in the previous time stage $t-1$. Error terms of past time points are used to predetermine future and present point's results. Moving average (MA) get rid of non-determinism or non-uniform movements from a time series. The property Q depicts Moving Mean (Average) in ARIMA. It is expressed as $MA(x)$ where x depicts past observations that are utilized to calculate present observation. Moving average models have a definite window and weights are related to the time. This means that the MA models are extra responsive to present event and are more reactive. MA is an abbreviation for moving average, which is also known as the rolling mean. Fundamentally, we are considering the simple average in each time frame and dividing it by the total number of time frames taken. Whereas Integration denotes the number of times the variations are taken. If a trend is there, then the time series is taken as non-stationary and depicts seasonality. Integrated is a quality which reduces seasonality from a time series. ARIMA models have differencing degree which removes seasonality.

4 SIMULATIONS AND RESULT

4.1 Case Study 1

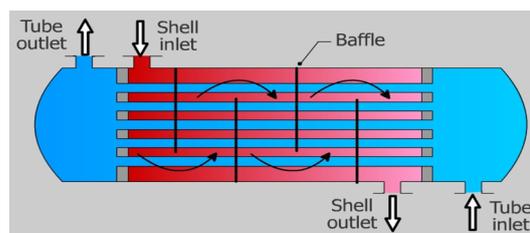


Figure 3. Shell and Tube Heat Exchanger

Shell and Tube type Heat Exchangers are popular types of heat exchanger because of design flexibility and wide range of temperatures and pressures. It consists of number of tubes mounted inside a shell as shown in the Figure 3. Two fluids or gases can exchange heat, one flows through the tubes and another from the outside of the tubes. The heat exchangers are generally manufactured according Tubular Exchanger Manufacturers Association (TEMA). As fluids passes through the shell and tube side during their regular operations, after a specific period of time the dirt or unwanted substances from the fluids gets settled on the surfaces. This deposition of unwanted material is called as Fouling and the theoretical resistance which represents the fouling is called as Fouling Factor. Hence maintenance is an important activity

in heat exchangers to prevent fouling. This maintenance can be carried out using predictive maintenance. The first step for any predictive modelling is a data. A sensor data set of heat exchanger consisting of five months data was obtained. Now once we receive data, it has to be classified whether it is structured or unstructured, continuous or logistic (Binary). Further the dataset is checked for null values, if any than they are eliminated. The dataset should now be classified for dependent and independent variable. Now, our data is ready for visualization. Data can be visualized using various tools like excel, tableau or python. The trend component in the data visualized in our case study, i.e. fouling factor is the dependent variable which is to be predicted. Then using multi linear regression machine learning algorithm the fouling factor is predicted using independent variables by training the model and testing. In this case study regression is used because the variable to be predicted i.e., fouling factor is completely dependent on independent variables. Table 1 depicts the Primary Variables i.e. measured variables and Secondary Variables i.e. calculated variables. The Independent variables in our case where Log Mean Temperature difference (LMTD), U-Transfer Rate and Q which is amount of heat transfer which are responsible for predicting fouling. Fig. 4. and Figure 5. depicts the relation between the U Transfer Rate and Cummulative Flow Tonnes per day with Fouling Resistance. The multi linear regression uses simple mathematical equation of $Y = mX + C$. Here Y is our dependent variable and X is independent variable. Using this model, the accuracy of the prediction came near about 97 percent and as shown in Figure 9. The visualization of test vs predicted data is shown in Figure 6. Also, for the same case study the fouling factor for future days was predicted using Time Series Forecasting as shown in Figure 7. and Figure 8. This time series forecast was predicted using Auto Regressive and Integrated Moving Average (ARIMA) model. Hence with the help of predictive maintenance using sensor data the fouling factor was predicted in order to get the trends and schedule the maintenance activity in advance. Also using ARIMA model the useful life was predicted till which the heat exchanger can continue its operations within the limits.

Table 1. Variable List

Primary Variables		Secondary Variables	
Hot Temp In	Hot Temp Out	Hot Temp Increase	Q
Cold Temp In	Cold Fluid Temp Out	Cold Temp Decrease	Hot-Out Cold-In
Cold Flow Rate	Hot Flow Rate	Fouling Resistance	Hot-In Cold-Out
Furnance Inlet Temp		U Transfer Rate	CMTD
		LMTD	

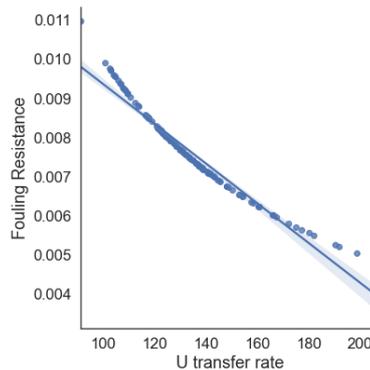


Figure 4. U Transfer Rate vs Fouling Resistance

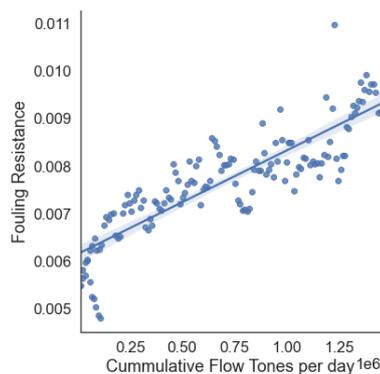


Figure 5. Cummulative Flow Tonnes Per Day vs Fouling Resistance

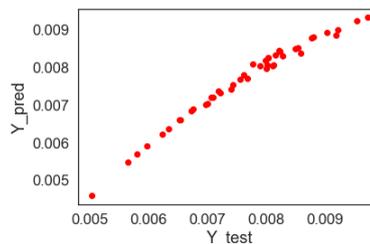


Figure 6. Fouling factor prediction of shell and tube heat exchanger

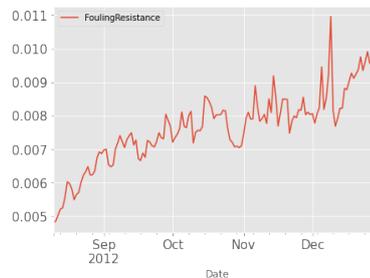


Figure 7. Working months vs Fouling Resistance

4.2 Case Study 2

Plate type heat exchanger is a specialized design to transfer heat between medium and low-pressure fluids. It is

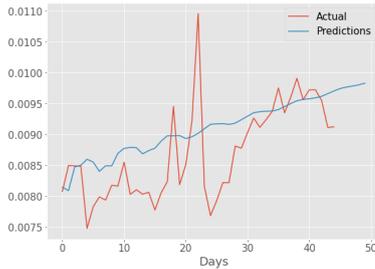


Figure 8. Actual vs Predicted fouling values

Intercept: 0.013079680465101827
 Coefficient
 U transfer rate -4.296929e-05
 Cumulative Flow Tones per day 4.609254e-10
 R2 Score is: 0.9741388398089696

Figure 9. R2 Score for shell and tube heat exchanger

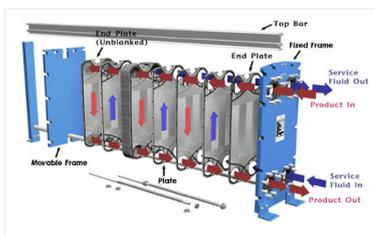


Figure 10. Plate Heat Exchanger

found to be most efficient of all heat exchangers. Plate type heat exchanger consist of a series of parallel plates as shown in the Figure 10. that are placed one above the other to allow the formation of series channels for fluids to flow between them. These heat exchangers are mainly used in oil, gas and chemical industries, food industries because they are exceptionally durable and ideal for transferring corrosive fluids or fluids which have high temperature. Fluid flows through plate type heat exchanger for transfer of heat for further processes during this process after some period there is accumulation of dirt inside or outside the tubes which forms small coating that adds resistance to heat transfer know as fouling factor. So, maintenance is an important activity to avoid fouling resistance in heat exchanger. This can be done using predictive maintenance. Initial requirement is data with which we can predict the future outcome. Now once we receive data, it must be classified whether it is structured or unstructured, linear, or logistic (Binary). Further the dataset is checked for null values, if any than they are eliminated. The dataset should now be classified for dependent and independent variable. our data is ready for visualization. Data can be visualized using various tools like excel, tableau or python with this Data Trend is visualized. Then using multi linear regression machine learning algorithm the fouling factor is predicted using independent variable by training and testing model. Fouling factor is completely dependent on independent variable so there is requirement to use regression. Independent variables in this case study where Log

Mean Temperature difference (LMTD), U-Transfer Rate and Q which is amount of heat transfer which are responsible for predicting fouling. Table 2 depicts the primary Variables i.e. measured variables and secondary Variables i.e. calculated variables. Linear Regression equation is used to show the dependence of one variable on another variable. As fouling resistance is dependent variable in both case studies 1 and 2, all the visual representations of this case study are similar to that of Case Study-1 i.e. Linear in nature. This model gave us accuracy and predicted value as 82 percent as shown in figure 11. Hence using predictive modelling, we can schedule maintenance of equipment well before its failure.

Table 2. Variable List

Primary Variables		Secondary Variables	
Thickness	Fluid 1	Q	Qr
Temp1 In	Fluid 2	Ae	LMTD
Temp1 Out	dP1	V1 In	V1 Out
Temp2 In	dP2	m1	m2
Temp2 Out	Area	Theta 1	Theta 2
		U	F
		A	M
		Me	NP

Intercept: 7.238990730530886e-06
 Coefficient
 Lmtd 9.247570e-05
 U 9.149892e-10
 Q -1.470825e-07
 Qr 2.328605e-08
 Me 6.552722e-06
 M -5.752454e-06
 R2 Score is: 0.8235290396306056

Figure 11. R2 Score for plate heat exchanger

4.3 Case Study 3

This set of data is created as per the semiconductor industry and contains sensor accounts from high exactness and innovative creation gear. Essentially, the semiconductor creation comprises of many cycle steps performing physical and substance procedure on so called wafers, for example cuts dependent on semiconductor material. Normally, lots of wafers are totalled into purported loads of size 25, which consistently go through similar tasks in the creation chain. This informational index was created as per the semiconductor business and contains sensor accounts from high-exactness and cutting-edge creation gear. Essentially, the semiconductor creation comprises of many cycle steps performing physical also, substance procedure on purported wafers, for example cuts based on semiconductor material Regularly, lots of wafers are collected into supposed bunches, which consistently passthrough similar tasks in the creation chain. In this chain, every process component is equipped with different sensors, for recording various parameters such as gas flow, temperature, volt-

ages etc. resulting in the recording of the data at each step of the process. Using a time frame in each one of the sensor data recorded and found important for the process by the experts, key numbers (KNs) among the dependent sensor data are recorded. The key numbers are used to interrupt in case of miss direction for keeping the whole process balanced. After the creation, every gadget on the wafer is tried in the most cautious route coming about in purported wafer test information. In a few cases, dubious examples happen in the wafer test data potentially prompting disappointment. For this situation, the underlying driver must be found in the creation chain. So, the key numbers were provided for the purpose. And the purpose was to find the main cause by finding the correlations in between the key numbers and the wafer test data. Henceforth during Exploratory Information Analysis (EDA) it was tracked down that out of 50 KNs as it were some KNs were showing varieties. Hence forth these sensors were utilized to anticipate the class of disappointment whether 'Good' or on the other hand 'Bad' as visualized in Figure 12. and Figure 13. There is a threshold value specified i.e., 0.75. If response goes beyond it then the classification will be bad otherwise good. If the response is 'Good' it will be used, otherwise it will be discarded. The expect action calculations utilized for this situation study was Support Vector Machine (SVM) and Principal Component Analysis (PCA). SVM is a type of supervised machine learning algorithm that can be used to solve classification or regression problems. Principal Component Analysis (PCA) is an unsupervised, nonparametric statistical technique basically used for dimensionality reduction in machine learning. The precision drew close to around 80 percent from Figure 14.

Hence by using previous data we can train the model for the "Good" and "Bad" response and therefore it can be used in real time on the field with proper testing to automatically classify the faulty wafers in the industry.



Figure 12. Training set of Good vs Bad data

5 CONCLUSION AND FUTURE SCOPE

The methods undertaken in this work are industry intensive. They help us obtain fairly accurate inferences about the state of the system. It is possible that instrument of measurement improve with time but not the method of performing analysis, diagnosis and repair, as they have remained more or less the same over many years. The

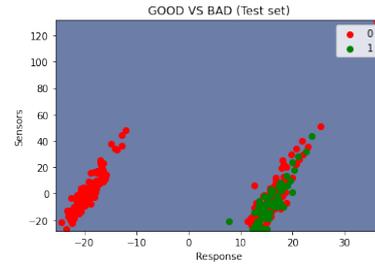


Figure 13. Testing set of Good vs Bad data

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(Y_test, Y_pred)
accuracy = float(cm.diagonal().sum())/len(Y_test)
print("\nAccuracy Of SVM For The Given Dataset : ", accuracy)
```

Accuracy Of SVM For The Given Dataset : 0.8

Figure 14. Accuracy of svm algorithm for semiconductor manufacturing process

methodology and technology is an exact science, but engineering isn't hence quality of work depends upon the skill and experience of the performer of monitoring and maintenance operation. More conformity in quality of reading as the data transmission is automated. Hence through this work, we predicted the fouling resistance of heat exchanger. The model was trained using machine learning and hence the fouling resistance is predicted using testing dataset which came out with prediction accuracy of approximately 97 percent. The prediction was done using R-squared technique which is calculated by the difference between the testing dataset and predicted value. The value should be very much nearer to 1. The more the R score nearer to 1 the more accurate will be the predicted value. Also done the time series forecasting using ARIMA model to predict the Remaining Useful Life(RUL) of process equipment. In the manufacturing wafer industry the faulty sensors are identified and the class is predicted using SVM and PCA. Hence there are tremendous opportunities in variety of industries with the use of huge amount of data to study the behaviour of the process instruments and for the predictive analytics and monitoring. Proper domain knowledge and analytical skills will support the digital transformations taking place and lead to a better future. The introduction of reliable cloud platforms used in condition monitoring, the development trend towards predictive maintenance, and strong demand from new applications are expected to drive the growth of the system condition monitoring.

References

- [1] P. Roback, J. Legler, Review of Multiple Linear Regression, in *Beyond Multiple Linear Regression* (Chapman and Hall/CRC, 2021), pp. 1–38, <https://doi.org/10.1201%2F9780429066665-1>

- [2] A. Gelman, J. Hill, A. Vektari, Logistic regression, *Regression and other stories* (Cambridge University Press, 2020)
- [3] D. Shaub, Fast and accurate yearly time series forecasting with forecast combinations, *International Journal of Forecasting* **36**, 116 (2020)
- [4] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, A. Lopez, A comprehensive survey on support vector machine classification: Applications, challenges and trends, *Neurocomputing* **408**, 189 (2020)
- [5] L. Silvestri, A. Forcina, V. Introna, A. Santolamazza, V. Cesarotti, Maintenance transformation through Industry 4.0 technologies: A systematic literature review, *Computers in Industry* **123**, 103335 (2020)
- [6] H. Truong-Ba, M.E. Cholette, P. Borghesani, L. Ma, G. Kent, Condition-based inspection policies for boiler heat exchangers, *European Journal of Operational Research* **291**, 232 (2021)
- [7] T. Zonta, C.A. da Costa, R. da Rosa Righi, M.J. de Lima, E.S. da Trindade, G.P. Li, Predictive maintenance in the Industry 4.0: A systematic literature review, *Computers & Industrial Engineering* p. 106889 (2020)
- [8] A. Jimenez-Cortadi, I. Irigoien, F. Boto, B. Sierra, G. Rodriguez, Predictive maintenance on the machining process and machine tool, *Applied Sciences* **10**, 224 (2020)
- [9] T.P. Carvalho, F.A. Soares, R. Vita, R.d.P. Francisco, J.P. Basto, S.G. Alcalá, A systematic literature review of machine learning methods applied to predictive maintenance, *Computers & Industrial Engineering* **137**, 106024 (2019)
- [10] M. Compare, P. Baraldi, E. Zio, Challenges to IoT-enabled predictive maintenance for industry 4.0, *IEEE Internet of Things Journal* **7**, 4585 (2019)
- [11] S.M. Lee, D. Lee, Y.S. Kim, The quality management ecosystem for predictive maintenance in the Industry 4.0 era, *International Journal of Quality Innovation* **5**, 1 (2019)
- [12] V. Radhakrishnan, M. Ramasamy, H. Zabiri, V. Do Thanh, N. Tahir, H. Mukhtar, M. Hamdi, N. Ramli, Heat exchanger fouling model and preventive maintenance scheduling tool, *Applied Thermal Engineering* **27**, 2791 (2007)
- [13] H.M. Hashemian, State-of-the-art predictive maintenance techniques, *IEEE Transactions on Instrumentation and measurement* **60**, 226 (2010)
- [14] S. Selcuk, Predictive maintenance, its implementation and latest trends, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* **231**, 1670 (2017)
- [15] Z. Zhou, B. Yao, W. Xu, L. Wang, Condition monitoring towards energy-efficient manufacturing: a review, *The International Journal of Advanced Manufacturing Technology* **91**, 3395 (2017)
- [16] R. Sahal, J.G. Breslin, M.I. Ali, Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case, *Journal of manufacturing systems* **54**, 138 (2020)
- [17] G. James, D. Witten, T. Hastie, R. Tibshirani, *An introduction to statistical learning*, Vol. 112 (Springer, 2013)
- [18] Position Paper- Deloitte Analytics Institute, <https://www2.deloitte.com/content/dam/Deloitte/de/Documents/deloitte-analytics/> (2017)