INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 4 - V7I4-1638)
Available online at: https://www.ijariit.com

Identifying defects during semiconductor manufacturing using Machine Learning

Nandini G.
nanduamma@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

K. G. Ashwin Krishnan
ashwinkrishna85@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Harshith R.
rharshith1868@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Dileep Kumar Simhadri
dileep.simhadri77@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Gummalla Akhil Kumar Reddy
akhil851982@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

ABSTRACT

Semiconductor production is the most technically advanced or difficult operations in the world. Old approaches of machine learning techniques, such as univariate and multifactor analysis, have been used to create prediction models for fault detection. In the last ten years, large joint research efforts between semiconductor manufacturing industry and academics have been launched in the field of fabrication. We explore some of these study topics in this paper and present techniques of machine learning which are used for automatic generation of a predictive model of maximum accuracy to detect defects throughout the semiconductor industry's wafer fabrication process. This research work attempts to develop a decision model that will assist in recognizing any equipment failures as fast as possible in order to maintain high productivity in manufacturing of semiconductors.

Keywords— *Semiconductor, Manufacturing, Equipment's, Delicate, Machine Learning, Algorithms*

1. INTRODUCTION

One of the most advanced and capital-intensive commercial areas are semiconductors manufacturing. Effective defect detection prediction in equipment is required to avoid equipment failures, as well as to increase production, lower costs, and reduce maintenance time. The creation of systems that allow computers to adjust their behavior depending on empirical data is known as machine learning (ML). ML analyses data and using statistical theory to create mathematical models that predict future events. Machine learning approaches are increasingly used in a range of manufacturing and scientific settings, involving technology-intensive production and, more broadly, any field which is data-intensive that could benefit accurate prediction proficiency, like the semiconductor manufacturing industry Manufacturing of semiconductors includes compound procedures. The number of stages in wafer manufacturing, which is often over 500, along with the amount of data gathered throughout the whole manufacturing process results in a massive amount of data which needs to be monitored .

The major manufacturing procedure in semiconductor manufacturing is: Manufacture of integrated circuits onto raw bare silicon wafers ,assembly of the integrated circuit into a package to make a ready-to-use product, and testing of the finished products. Majority of semiconductor production or manufacturing machines are fitted with detectors or sensors in over the last few years to enable the monitoring of the production process in real time. These sensor data from the production and equipment states allow for more optimization and efficient control. Sadly, such data which is measured is so profuse that it is difficult to come across any issue throughout the manufacturing process in a well-timed manner. The topic of reliable identification of failure of the manufacturing equipment in the wafer production process is investigated in this work. Machine learning techniques can be used to automatically create a fault detection model from existing sensor data. The hunt for an effective and efficient method to keep an eye on the equipment health and detect at hand breakdown and has long piqued of both researchers and industry.

2. RELATED WORK

Cost, quality, and delivery time are all important variables in most manufacturing processes if businesses want to compete in the long run. Process engineers must detect or monitor and find the unique specifications/characteristics of aberrant goods as

soon as feasible during the manufacturing process. Process control is critical in the semiconductor industry, which uses multi-level manufacturing processes to produce the outcome with a productscale of less than 300 nanometers.

The measurement data collected from the sensors placed in the equipment and the data received from the last electrical test, modern semiconductor manufacturing technology allows for real-time process management. With such a large number of data collected throughout the manufacturing process, process engineers face a difficult task in effectively monitoring and controlling the process by researching and evaluating the data. Traditional process control methodologies, such as univariate and multivariate control charts, are no longer effective for controlling the manufacturing process which includes many stages of processing. Rather, a modern and automated process control system is vital.

To detect faults in plasma tech equipment, Ison and colleagues suggested a decision tree classification model. The model was created using the data from the five sensor signals. During the etching process, many researchers looked into the topic of fault detection. Godlin et al advocated creating a custom chart for indicating problem types. They directly obtained tool-state data from the etcher. There are 19 variables in this set of data. The statistical strategy was also used in the work of Spitzlsperger and colleagues. By using the re modeling strategy, they were able to sustain variations in the standard deviation coefficients and mean by using the multivariate control chart method.

3. SYSTEM DESIGN ANDMETHODOLOGY

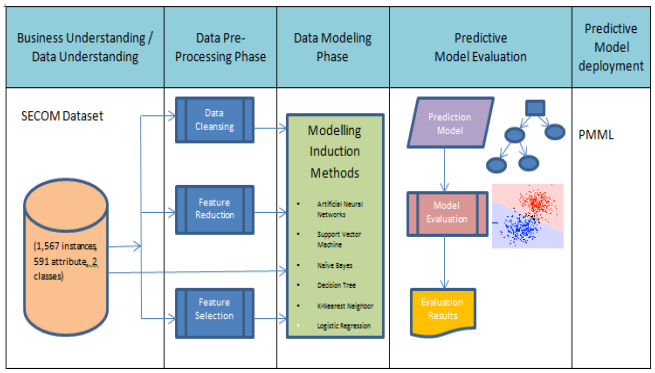


Figure 3.1 System Design

There are five phases in the project:

3.1 Data preparation phase

The first and most important part in constructing a prediction model is information gathering and preparing. Data preparation is a necessary step in converting multiple data and types into a form that is understandable by a machine learning predictive model. Larger volumes of data are acquired on a regular basis during the semiconductor manufacturing process. All factors, including predictor variables, are included in the collected data, which can be utilized to create prediction models. The data supplied is "horizontal": there is a large number of different variables (must be minimized) and just a few observations under the same operating conditions**.** To reduce the number of regressors with correlation analysis, and PCA,variable selection. While we want to raise the total observations useable for modeling with data clustering. The machine has collected over one hundred statistical variables like means, variances, maximum and minimum values, and so on**.**

3.2 Data pre-processing Phase

a) Data Cleansing Phase: In predictive modeling, the data purification phase is crucial. The SECOM data file must be processed for irregularities and the information must also be adjusted because the raw data'svalue ranges vary greatly. 452 instances with null and missing values are discarded as a result of a data cleansing exercise.

b) Feature scaling: The approach used to normalize the data set was called "feature scaling." The purpose of cleansing of data is to get a fully clean and complete data collection that can be matched or modeled and by removing outliers and missing data . The variables that are continuous are transformed on a linear scale to a specified value with a range of 0 to 1 or 0 to -1 to 1 and to normalize the input data set. Ordinal data were evenly spread across the same range. The class mean was used to fill in for missing values. In neural networks, data of diverse sizes can cause instability (Weigend and Gershenfeld, 1994). The following normalization equation is used to normalize the data which is the raw data of input and output:

$$x_{\text{norm}} = 2\times\frac{(x - x_{\text{min}})}{(x_{\text{man}} - x_{\text{min}})} - 1$$

Where *x* indicated the data which is to be normalized, i.e., and *x* min and *x* max are minimum and maximum values of the data which is raw.

c) Feature reduction: One of the most difficult aspects of the semiconductor production process is the large number of tool variables [2], which makes variable selection strategies particularly valuable[7]. Examine the data collected from sensor's, i.e., the data specified in each column. Remove the feature if the data appears to be a single or individual. Count the "not available" or "missing" values in each column.

d) Feature Selection: We try to find the most influential parameters. The best procedure to select the right parameters is to utilize a trial/error procedure.

3.3 Data modeling phase and predictive model deployment

After preprocessing the data using various data preprocessing phase the next step is to model this data by using various machine learning algorithms. We have implemented this by using 3 machine learning algorithms. Naïve bayes, decision tree and random forest from which we have selected one Machine learning algorithm which showed maximum accuracy and deployed a model by utilizing a right algorithm.

4. ALGORITHMS USED

A Naive Bayes classifier is a classifier which is probabilistic based that was first proposed in 1973 by Duda and Hart . The SECOM data file is an unbalanced datafile, and Nave Bayes is a popular solution for unbalanced dataset issues. In the realm of classification, the Naive Bayes induction algorithm is very desired and approved. The Naive Bayes Classifier approach basically works on the Bayesian theorem and is well-suited to the Semiconductor manufacturing sector, where input dimensionality is large. Naive Bayes can often outperform more complicated classification systems, despite their simplicity.

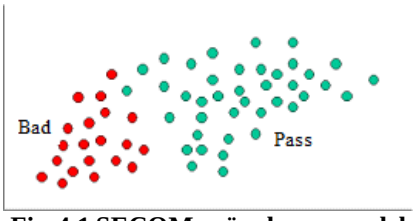


Fig-4.1 SECOM naïve bayes model

The model properly projected the pass class for preventive maintenance 44 times and mistakenly forecasted it 15 times in this confusion matrix. For predictive purposes, the model accurately predicted the negative class 144 times and mistakenly predicted it 265 times.The conditional probabilities are what the Nave Bayes method is based on the foundation of this method is Bayes' Theorem. It employs a formula that "calculates a probability based on the frequency of values and combinations of values in historical data." Itgoes like this:

Prob(B given A) = Prob(A and B)/Prob(A and B)

Decision tree: A important method for discovering a tree-based model for future based event prediction is decision tree induction. A decision tree is a tree like structure that reflects a collection of choices. Each non-terminal node indicates a test or judgement that will be performed on a singleattribute value (i.e., input variable value) of the evaluated data item, with one branch and sub-tree for each potential test outcome.

K Means classifier: One such study develops a prediction model employing the k-nearest neighbor method (PD- kNN) due to the particular properties of laser diodes, including such in homogeneity in the data. The k nearest rule is an easy-to-understand notion with the following central assumption: The kNN rule selects the k nearest labeled examples in the learning algorithm set and allocates x to the category that typically forms inside the k-subset for a particular unmarked input x. (i.e., k-nearest neighbors). The suggested prediction (PD-kNN) is premised on the reality that the gap between the faulty example and the nearest neighboring training data has to be bigger than that of the separation between the as any and the nearest neighboring target variables. The goal is to calculate a threshold (t) with a particular level of certainty. The approach is divided into two sections:

4.1 Development of Prediction Models

A KNN quadratic value is indeed the total of sampling is squares values to its k-nearest neighbors

4.2 Classification-Based Fault Diagnosis

Classification for fault detection

The fault diagnosis element of an inboundtraining dataset x involves three steps

1) As from learning database, identify x's k-nearest neighbors.
2) Calculation of D2x, which is x's kNN squareddistance
3) Correlation of D2x to the T-strength.

It is categorized as a standard if D2x = T; else, itis identified as a fault.

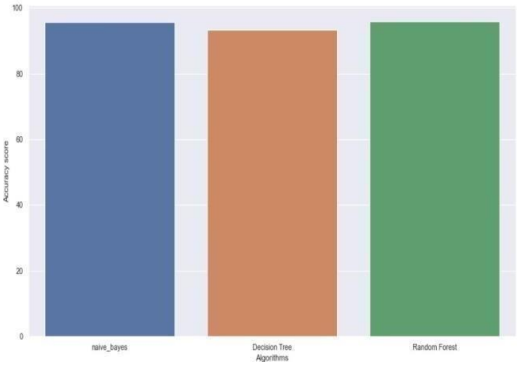
4.3 Random Forest

A random forest is a supervised classification system that uses decision trees to build it. This technique is used to anticipate processes and performance in a number of industries, and we have also implemented random forest algorithm in our project. A random forest is an algorithm for solving classification and regression. It makes use of supervised methods, which is really a tool to solve complicated problems by combining several classifiers. Several decision tree algorithm make up a random forest algorithm. The random forest algorithm's produced 'forest' is taught via bagged or bootstrapping aggregation. Packing is a schema that enhance the reliability of machine learning by grouping them together.

The goal of Random Forests is to lower the variability of CART's predicted value and increase the standard errors. The approach can also assess its very own effectiveness by testing the model's predictive accuracy and calculating accuracies utilizing information not chosen via bootstrapping (out-of-bag or OOB samples)(OOB error).

5. RESULTS

By using naïve bayes ,decision tree and random forest algorithm we have got a clear picture of which algorithm gives maximum accuracy in detecting faults in the manufacturing of semiconductors. From the bar graph it is clear that random forest has proven to give maximum accuracy of 96 percent. This methodology can be used in predictive maintenance.



6. CONCLUSION

The manufacturing of semiconductors requires high capital and high initial investment and also requiring a high level of investment on equipment's required for manufacturing. Manufacturing equipment and its optimizing has garnered a lot of consideration and it has shown to be very advantageous. Engineers and researchers have interesting problems and opportunitiesin developing a new quality grade for this rapidly expanding industry. A legible prediction and classification model is useful for predicting the fabrication process of semiconductors. Most semiconductor production is quite complicated, and hundreds of metrology data are constantly generated and waiting to be analyzed by process engineers for the goal of getting good yield in high quality product.

7. REFERENCES

[1] P. Zhao, M. Kurihara, J. Tanaka, T. Noda, S. Chikuma and T. Suzuki, "Advanced correlation-based anomaly detection method for predictive maintenance", *2017 IEEE International Conference on Prognostics and Health Management(ICPHM)*, pp. 78-83, 2017.

[2] B. Cline, R. S. Niculescu, D. Huffman and B. Deckel, "Predictive maintenance applications for machine learning", *2017 Annual Reliability and Maintainability Symposium (RAMS)*, pp. 1-7, 2017.

[3] J. S. L. Senanayaka, S. T. Kandukuri, Huynh Van Khang and K. G. Robbersmyr, "Early detection and classification of bearing faults using support vector machine algorithm", *2017 IEEE Workshop on Electrical Machines Design Control and Diagnosis (WEMDCD)*, pp. 250-255, 2017.

[4] T. S. Buda, H. Assem and L. Xu, "ADE: An ensemble approach for early Anomaly Detection", *2017 IFIP/IEEE Symposium onIntegrated Network and ServiceManagement (1M)*, pp. 442-448, 2017.

[5] J. Shimada and S. Sakajo, "A statistical approach to reduce failure facilities based on predictive maintenance," 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, 2016, pp. 5156-5160.

[6] Z. Yang, P. Baraldi and E. Zio, "A comparison between extreme learning machine and artificial neural network for remaining useful life prediction," 2016 Prognostics and System Health Management Conference (PHM-Chengdu), Chengdu, 2016, pp. 1-7.

