Micron Internship Final Presentation

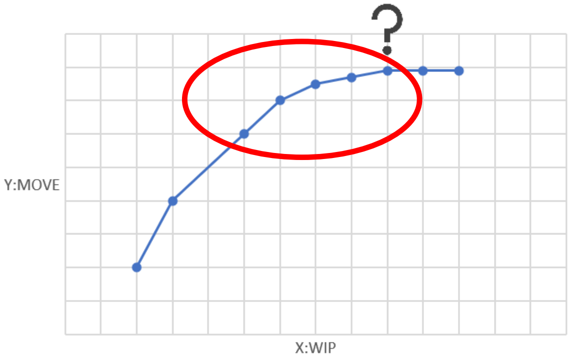
Application of Machine Learning and KNN Algorithm to Predict Ideal WIP

Position : F16 ME PROD shift team intern
Name : Cheng-Yan Yang
Updated Aug 2022

Contents

01. Introduction

- During wafer fabrication, process data, equipment data will be automatically recorded and accumulated in database for monitoring the process, diagnosing faults.
- However, in semiconductor manufacturing industry, many factors that are interrelated affect the yield of fabricated wafers.
- Therefore, this study uses Machine Learning methods to analyze when WIP raise to a certain level, MOVE will approach saturation.



01. Introduction

- **Purpose:** Hope to find ideal WIP through Machine Learning methods.
- **Methods:** Using KNN algorithm and data analysis process to analyze and predict ideal WIP.
- **Analysis Tools:** Python、SQL、Excel

- **Why choose KNN model?** (R squared, MSE...)
- **What are the analysis variables?** (WIP, MOVE)
- **How many data do we need in this research?** (One year data)

	R squared	MSE	Data type
K Nearest Neighbors	WIN 0.93	WIN 0.69	Nonlinear
Linear Regression	0.65	3.67	linear
Support Vector Regression	0.78	2.26	Nonlinear

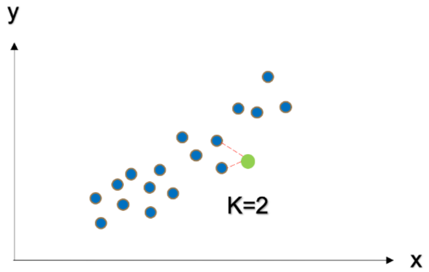
02. Literature Review - KNN algorithms

KK Nearest Neighbors(KNN)

It belongs to one of the algorithms of Supervised learning in Machine learning.

Pros : The biggest advantage of k-NN is that it is easy to understand, and the performance is reasonable without too much adjustment.

Cons : The construction of the k-NN model is fast, but the prediction speed is slow when the number of features or data points in the training data is large.



identifier	class name	args	distance function
"euclidean"	EuclideanDistance	•	`sqrt(sum((x - y)^2))`
"manhattan"	ManhattanDistance	•	`sum(|x - y|)`
"chebyshev"	ChebyshevDistance	•	`max(|x - y|)`
"minkowski"	MinkowskiDistance	p	`sum(|x - y|^p)^(1/p)`
"wminkowski"	WMinkowskiDistance	p, w	`sum(|w * (x - y)|^p)^(1/p)`
"seuclidean"	SEuclideanDistance	V	`sqrt(sum((x - y)^2 / V))`
"mahalanobis"	MahalanobisDistance	V or VI	`sqrt((x - y)' V^-1 (x - y))`

02. Literature Review - KNN algorithms

Evaluation Matrix ：MSE、R Square

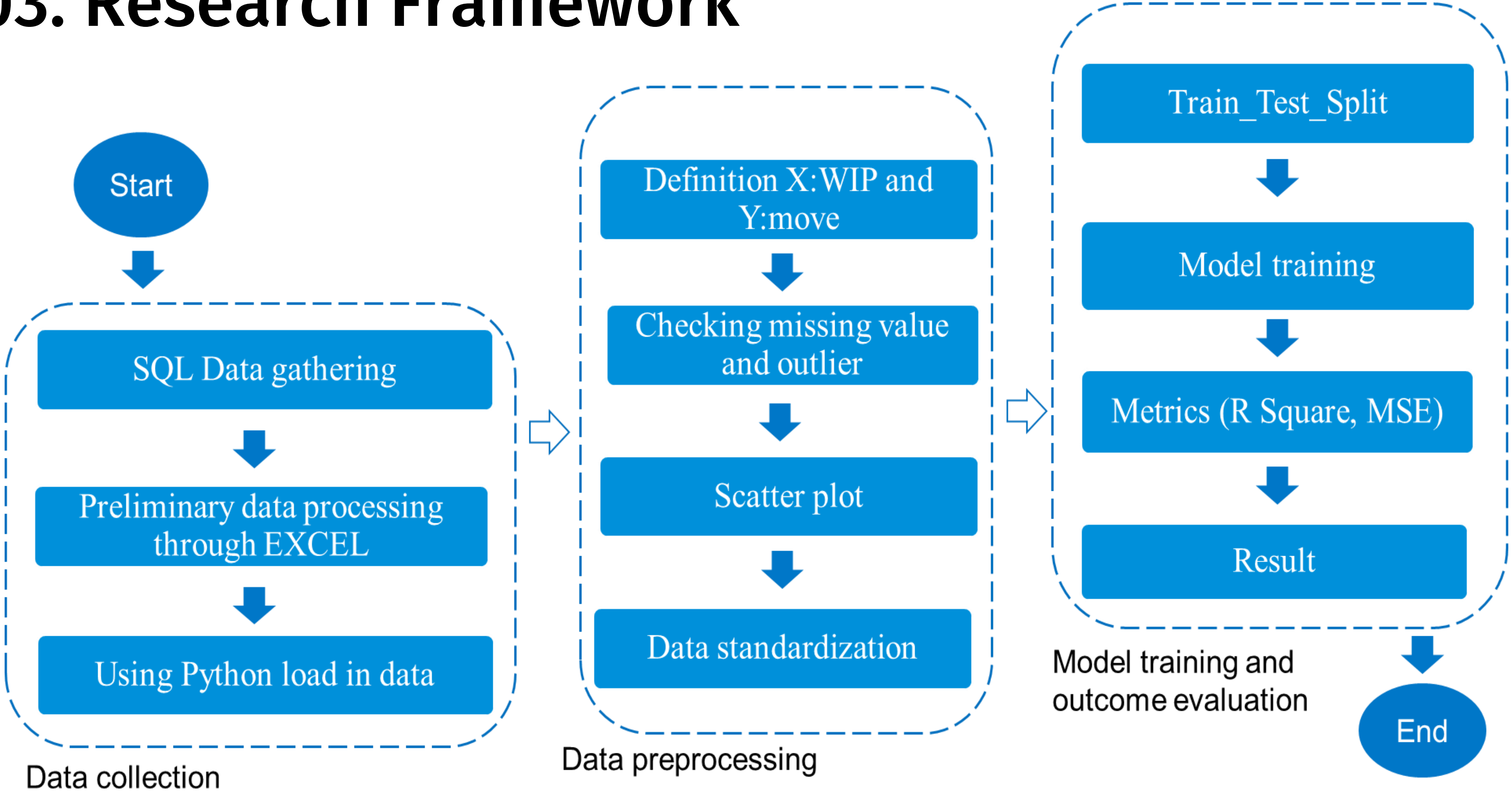
$$MSE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

MSE (Mean-Square Error) The smaller the error, the better。

$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

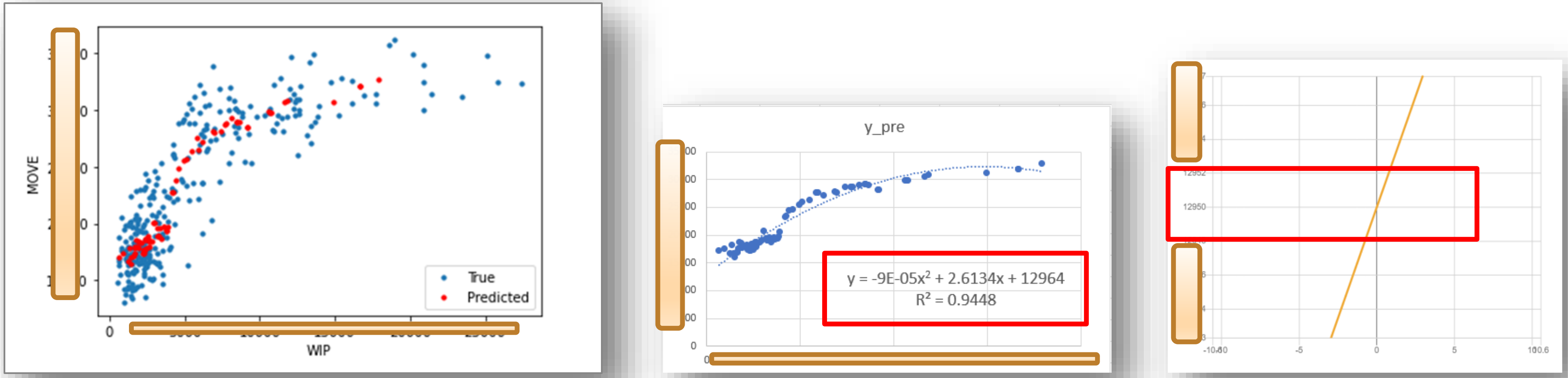
R Square (coefficient of determination):The coefficient is between 0 and 1, and it is used to explain the percentage of explained variation in the total variation. The closer to 1 is the better.
(R2 > 0.8 are highly reliable)

03. Research Framework



04. Experimental Results

This is one of the Dry Etch bottleneck work stations, as the research object of this time.

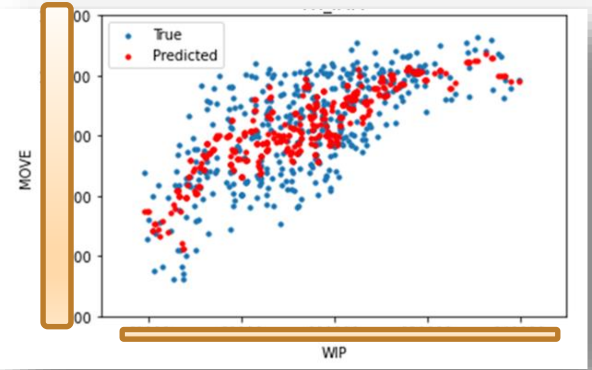


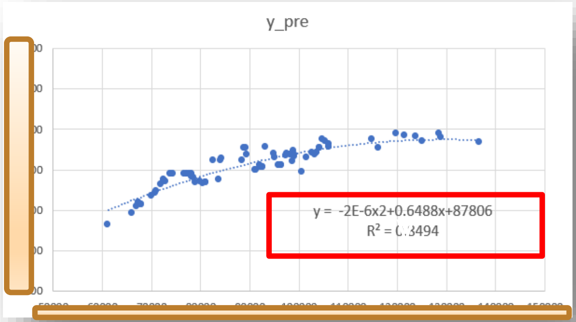
- For example, we chose Dry Etch. When the R Square of the KNN model is 94.48% and the MSE error is 0.011, we can know that the reasonable WIP is 12.95k. If the WIP exceeds 12.95k, the Move will approach saturation at 30k.

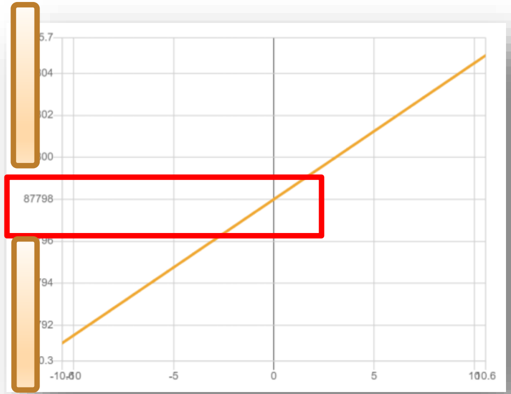
(MSE: 0.011 , R2: 0.94 , ideal WIP: 12.95k)

04. Experimental Results

This is one of the PHOTO bottleneck work stations, as the research object of this time.







- **In PHOTO bottleneck stations. When the R Square of the KNN model is 84.94% and the MSE error is 0.044, we can know that the reasonable WIP is 87.8k. If the WIP exceeds 87.8k, the Move will approach saturation at 90k.**

(MSE: 0.044 , R2: 0.8494 , ideal WIP: 87.8k)

04. Experimental Results

Conclusions and Future Research

AREA	WSG	Ideal WIP
PHOTO	XXXX	87.8K
Dry Etch	XXXX	12.95K

1. The concept is extended to each WS of other Area, and the same analysis method is used to find the ideal WIP.
2. Add other factor analysis (MA 、 MU 、 idle 、 lost) ,using (Principal components analysis, PCA) Achieve data dimensionality reduction and find key factors.

References

Math Solver : https://mathsolver.microsoft.com/zh-Hant

KNN回歸演算法 https://ithelp.ithome.com.tw/articles/10269826

資料分析流程 https://ithelp.ithome.com.tw/articles/10184905

機器學習 http://yltang.net/tutorial/dsml/12/

混淆矩陣 指標評估 https://jason-chen-1992.weebly.com/home/-confusion-matrix

https://medium.com/@s716419/%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92%E6%A8%A1%E5%9E%8B%E8%A9%95%E4%BC%B0%E6%8C%87%E6%A8%99-confusion-matrix-precision-and-recall-e9d64ff14d81

Chien, C. F., Hsu, C. Y., & Hsiao, C. W. (2012). Manufacturing intelligence to forecast and reduce semiconductor cycle time. *Journal of Intelligent Manufacturing*, *23*(6), 2281-2294..

Hsu, C. Y., Yang, C. S., Yu, L. C., Lin, C. F., Yao, H. H., Chen, D. Y., ... & Chang, P. C. (2015). Development of a cloud-based service framework for energy conservation in a sustainable intelligent transportation system. *International Journal of Production Economics*, *164*, 454-461.

Chien, C. F., Hsu, C. Y., & Chen, P. N. (2013). Semiconductor fault detection and classification for yield enhancement and manufacturing intelligence. *Flexible Services and Manufacturing Journal*, *25*(3), 367-388.