

摘要

本研究旨在利用工業 4.0 概念推動智慧製造的趨勢，透過資料探勘和機器學習技術，對半導體製程中的重要指標進行分析，利用網路上公開的資料集，其中來自半導體製造過程的 SECOM 資料集，以探討製程中的異常情況，為解決許多企業在資料探勘上的挑戰，本研究採用了簡易的機器學習平台 Azure Machine Learning Studio，透過該平台的拖拉模組的方式，使得中小企業也能輕鬆進行機器學習。

研究的目的主要在比較決策樹和隨機森林兩種機器學習模型在半導體製程中的效能，結果顯示，兩種模型在分類的過程中皆表現良好，都具有高準確性和預測能力，真陽性率(TPR)和真陰性率(TNR)高，偽陽性率(FPR)和偽陰性率(FNR)低，說明模型在識別不良品和良品方面的能力強，且誤報和漏檢的概率較低，然而，需注意到資料集中的負標籤(Negative Label)為不良品，兩個模型中偽陰性(FN)比偽陽性(FP)多，對科技業或製造業而言，可能導致重工或不良品流出等問題，且該成本是不可忽視的。

關鍵字：Azure Machine Learning Studio、機器學習、決策樹、隨機森林、資料探勘