

東 海 大 學

工業工程與經營資訊學系

碩士論文

機台維修保養時程規劃

-以半導體晶圓廠為例

研 究 生：徐子寒

指 導 教 授：王立志 教授

中 華 民 國 一 〇 七 年 七 月

東海大學

工業工程與經營資訊學系

碩士學位論文口試委員會審定書

本系 徐子寒 君

所提論文 機台維修保養時程規劃-以半導體晶圓廠為例

合於碩士資格水準，業經本委員會評審通過，特此證明。

口試委員：

袁明鑑

王宗新

王元志

指導教授：

王元志

系主任：

考欽印

中華民國 一〇七 年 七 月 六 日

東海大學

工業工程與經營資訊學系

碩士學位論文指導教授推薦書

本系 徐子寒 君

所提論文 機台維修保養時程規劃-以半導體晶圓廠  
為例

係由本人指導撰述，同意提付審查。

此致

工業工程與經營資訊學系

指導教授



日期 107年 7 月 30 日

東 海 大 學

工業工程與經營資訊學系

碩士論文

機台維修保養時程規劃

-以半導體晶圓廠為例

研究生：徐子寒

指導教授：王立志 教授

中華民國一〇七年七月

# **A Study on Machine Maintenance Schedule for Semiconductor Wafer Fabrication**

By  
Tzu-Han Hsu

Advisor : Prof. Li-Chih Wang

A Thesis  
Submitted to the Institute of Industrial Engineering and Enterprise  
Information at Tunghai University  
in Partial Fulfillment of the Requirements  
for the Degree of Master of Science  
in  
Industrial Engineering and Enterprise Information

July 2018  
Taichung , Taiwan

# 機台維修保養時程規劃-以半導體晶圓廠為例

學生：徐子寒

指導教授：王立志 教授

東海大學工業工程與經營資訊學系

## 摘 要

在全球科技技術迅速提升、電子產品快速成長，人工智慧和大數據技術已逐漸走向商業化的道路，在這樣的背景下，半導體扮演著至關重要的角色，而台灣又是在國際半導體產業位居前三名，因此如何有效幫助半導體產業降低製造成本、增加產能是半導體相關企業需要持續攻克的難題。以台灣半導體晶圓廠為例，生產線可能因為突發當機而影響整條產線運作，需要重新開機、維修調整等造成損失，如何維持機台正常運作，是確保產品良率的關鍵。

有鑑於此，本研究利用決策樹作機台當機之預測模型並結合模擬系統進行排程，根據歷史數據排出有效的維護計劃，以往工廠對於維修保養採用定量、定時保養；相反，本研究採用決策樹和模擬系統來開發維修決策支援模組，將上述列入決策樹投入因子作為探討，由企業資源規劃(ERP)、製造執行系統(MES)讀取其資料，以動態生成有效的維護計劃。將影響當機之因子進行決策樹分類，依據決策樹歸納出其法則，將決策法則參數設定並進行模擬排程，將其法則列入機台維修保養時程規劃中。本研究所發展維修保養決策樹方法，應用於半導體晶圓廠案例，可以發現利用維修決策支援模組能準確預測率高達 60%，當機比率明顯降低。與原始數據相比，修決策支援模組生成的維護計劃將導致較低的機器故障率 43%，若預先規劃保養、維修機台時程，能有效減少無預警造成的損失，可使機台稼動率提升。

**關鍵詞：** 半導體晶圓製造、維修保養、決策樹

# **A Study on Machine Maintenance Schedule for Semiconductor Wafer Fabrication**

Student : Tzu-Han Hsu

Advisor : Prof. Li-Chih Wang

Department of Industrial Engineering and Enterprise Information  
Tunghai University

## **ABSTRACT**

In the context of the rapid advancement of global technology and electronic products, Artificial Intelligence (AI) and Big Data have gradually turned to be commercialized. Semiconductor technology has become one of the most important technologies in this era and Taiwan ranked as the top three in the semiconductor industry of the world. Reducing manufacturing cost and increasing the production capacity are crucial for any semiconductor company due to the vast facility investment. Taking semiconductor plant in Taiwan for example, the throughput will be significantly affected by unexpected machine breakdown and scheduled maintenance, the right timing for scheduling machine maintenance is critical for ensuring production quality and reducing machine breakdown.

Consequently, this research applies the decision tree as the predictive model for the machine breakdown combined with simulation system for generating an effective maintenance schedule based on the historical production data. In the past, the factory usually adopted the predetermined fixed-quantity maintenance approaches such as time-based maintenance (TBM) and predictive maintenance of capacity (PMC). On the contrary, this study employed the decision tree and capacity planning simulation (CPS) to develop a maintenance schedule decision support model (MSDSM), which will integrate with ERP and MES, to dynamically generate an effective maintenance schedule. The maintenance decision tree obtained from MSDSM was evaluated by varying production quantity, maintenance time horizon and the machine numbers, results show that the prediction accuracy of machine breakdown can reach up to 60%. The MSDSM is also applied in a full-scale 12" semiconductor wafer fab, the maintenance schedule generated from MSDSM will result in a lower machine breakdown rate 43% and higher throughput compared to the original production data.

**Keywords: Semiconductor Wafer Fabrication, Maintenance, Decision Tree**

## 致謝

首先感謝這兩年來辛苦細心指導我的教授王立志博士，讓我有機會在 AUTO 學習，不管是在研究上、做事的態度上及實務經驗，帶領我往正確的方向，並且給予最大的鼓勵使我能順利的完成此篇論文。

在AUTO的兩年生活，除了大家一起熬夜趕報告，過著一天五餐的日子，挑戰我三天沒睡覺的極限紀錄；大家一起整人慶生、定期舉辦大胃王比賽、一起出國旅遊，研究室能一起認真做事，也能一起開心玩樂。因為有你們，讓我的研生活變得絢麗多彩。一度覺得自己無法順利完成論文，而現在我無法想像真正的完成論文的感覺，這兩年的酸甜苦辣即將畫上了句號，慶幸一路上遇到你們，因為有你們，我才能完成此篇論文，謝謝鳳芷與我分享研究所經驗談；謝謝莉莉、碧玲在這艱辛的道路上，陪伴著我、支持的我；謝謝阿邦在我最低潮時，不斷鼓勵著我，還得忍受我的脾氣；謝謝吐司怪獸當我的垃圾桶，聽我每天抱怨，在那邊的時光，總是讓我很放鬆；謝謝我身邊每一個人總是給予我幫助、鼓勵，你/妳們的幫忙及支持我銘感在心。

最後，謝謝我親愛的家人們，在這麼多年的求學過程中，一直忙於學業，與家人相處時間很少，但有了你們的支持與鼓勵，讓我無後顧之憂，順利完成學業。謝謝我親愛的妹妹，雖然我們平時愛鬥嘴，但在我低潮時，一直陪伴我、逗我開心，終於我們倆都順利畢業了。謹以此文獻給我摯愛的雙親。

子寒 于東海大學工業工程與經營資訊學系 2018年7月

# 目錄

摘要.....	i
ABSTRACT.....	ii
致謝.....	ii
目錄.....	iv
表目錄.....	v
圖目錄.....	vii
第一章 緒論.....	1
1.1 研究背景與動機.....	1
1.2 研究動機.....	4
1.3 研究目的.....	4
1.4 研究方法及步驟.....	5
第二章 文獻探討.....	7
2.1 維修保養策略.....	7
2.2 資料探勘.....	9
2.3 資料探勘應用於半導體.....	10
2.4 決策樹.....	13
第三章 維修決策支援模組.....	15
3.1 規則生成階段.....	17
3.2 模擬結果階段.....	31
3.3 績效比較階段.....	39
第四章 實證研究.....	40
4.1 規則生成階段.....	40
4.2 模擬結果階段.....	47
4.3 績效比較階段.....	49
第五章 結論與建議.....	53
5.1 結論.....	53
5.2 未來研究方向.....	54
參考文獻.....	55

## 表目錄

表 2.1 Data Mining 於半導體應用 .....	12
表 2.2 決策樹比較.....	13
表 3.1 原始資料整合.....	18
表 3.2 欄位說明.....	19
表 3.3 資料清潔.....	20
表 3.4 轉換欄位說明.....	21
表 3.5 資料轉換.....	22
表 3.6 模式範例法則.....	29
表 3.7 決策樹維修策略表.....	30
表 3.8 品項生產資訊表單.....	33
表 3.9 機台群組.....	33
表 3.10 訂單資訊.....	33
表 3.11 客戶訂單.....	34
表 3.12 產品製程表.....	35
表 3.13 製程時間表.....	35
表 3.14 機台群組表.....	36
表 3.15 機台開關機時間設定表.....	36
表 3.16 生產紀錄表.....	37
表 3.17 機台資訊紀錄表.....	37
表 3.18 機台開關機設定表.....	38
表 3.19 新增決策樹當機法則之機台資訊紀錄表.....	38
表 3.20 決策樹維修策略表-使用者設定 .....	39
表 4.1 機台維修.....	40
表 4.2 機台生產狀況.....	41
表 4.3 投入資料.....	41
表 4.4 原始資料表.....	43
表 4.5 參數設計.....	44
表 4.6 生產數量為 500 情況下.....	44
表 4.7 生產數量為 1000 情況下.....	45

表 4.8 關鍵機台當機法則.....	46
表 4.9 案例決策樹維修策略表.....	47
表 4.10 歷史資料機台當機.....	48
表 4.11 維修支援模組關鍵機台當機.....	48
表 4.12 IHC01 突發當機比率.....	49
表 4.13 IHC06 突發當機比率.....	50
表 4.14 機台 Down 比率總和比較.....	52

# 圖目錄

圖 1.1 台灣半導體產業營收趨勢.....	1
圖 1.2 半導體產業供應鏈結構.....	1
圖 1.3 浴盆曲線.....	3
圖 1.4 本研究之架構與流程.....	6
圖 2.1 資料採礦流程.....	10
圖 2.2 Weka 介面.....	14
圖 3.1 維護決策支援模組流程圖.....	16
圖 3.2 時間-數量關係圖.....	20
圖 3.3 分群法.....	24
圖 3.4 C4.5 流程.....	26
圖 3.5 範例決策樹.....	28
圖 3.6 決策樹正確率.....	29
圖 3.7 CPS 系統架構.....	31
圖 4.1 關鍵機台決策樹.....	46
圖 4.2 關鍵機台決策樹正確率.....	47
圖 4.2 IHC01 當機波動.....	51
圖 4.3 IHC06 當機波動.....	51

# 第一章 緒論

## 1.1 研究背景與動機

### 1.1.1 研究背景

半導體是影響國內經濟相當重要的一個產業，全球經濟環境成長的趨勢，帶動國際半導體市場的需求。台灣半導體產業已成熟發展，在總體營收方面持續成長中，前景值得期待。台灣半導體產業營收如圖 1.1，整體產業營收持續成長。

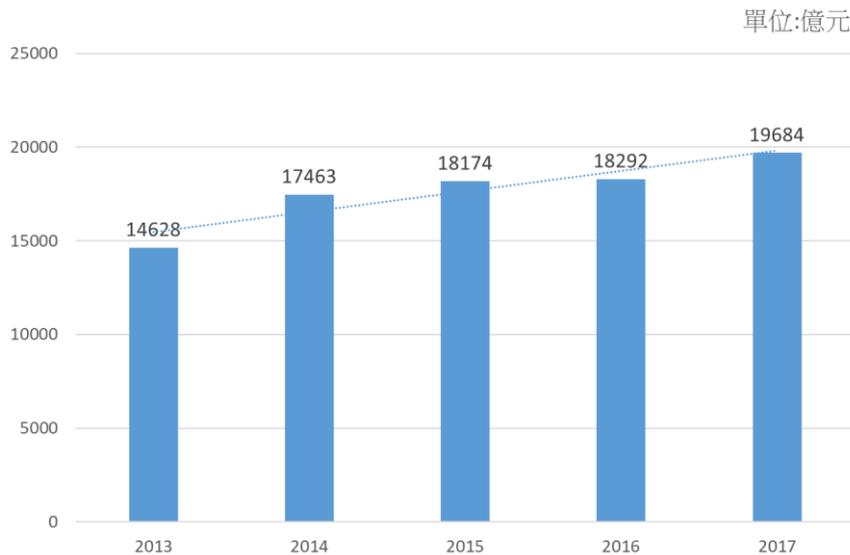


圖 1.1 台灣半導體產業營收趨勢

(資料來源：財務部資料中心)

半導體產業已是台灣製造業中重要的指標產業，其代工模式成熟，並且營業收入持續成長，全球產值排名第三。在台灣已經有一套完整的半導體產業的供應鏈結構，主要可以分成四大階段，積體電路設計、晶圓製造、封裝、測試，封裝業、測試業。

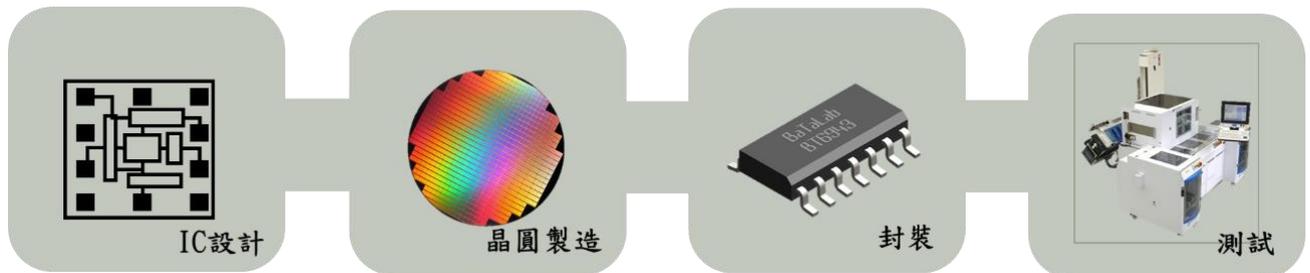


圖 1.2 半導體產業供應鏈結構

2017 年台灣半導體產業鏈的產值結構中，晶圓代工占 49%、IC 設計占 25.1%、IC 封測占 19.4%，總產值達 810 億美元，僅次於美國、韓國，全球排名第三，產值占全國 GDP 的 15%。其中又以晶圓代工業全球排名第一，市占率最高，占七成以上，產值達 397 億美元，由此可見晶圓製造業對於整體經濟扮演著重要的角色。全球製造業競爭力大，技術快速進步，如何提高毛利率，爭取價格上競爭優勢，使企業持續成長，為企業所需思考，半導體的製程複雜，影響產出的因子很多，除了維持生產良率外，如何保持機台能持續運作更是首要的條件，一旦突發當機影響整條產線運作，可能需要重新開機、維修調整，甚至花費過多時間維修機台，這些對於半導體廠均為損失來源，若能預測機台的健康狀態，預先規劃保養、維修機台，減少無預警造成的損失，避免突發當機，進而提升生產效率。

目前晶圓製造廠設備維修仍是設定為固定保養時間，各廠均訂有不同之標準，而保養依賴資深維修人員的經歷，但設備更新快速、時間限制、人員流動等因素，造成技術傳承不易，當機台發生異常時，才會依照機台狀況進行故障診斷。過去傳統的維修方法為被動式維修，不實施定期檢查，直到機台發生異常，工程師才由機台狀態進行故障診斷，嚴重時維修天數可能耗費多日，而影響整體產能下滑，且產品的良品率也會因此而降低，造成工廠成本上升、交期受影響。

對於設備維修保養的發展，從資料蒐集、整合量測到決策保養，如何透過資料探勘的方法來協助公司進行機台當機維修保養時程規劃確保機台正常運作、降低停機時間，成為影響高科技產業重要議題之一。

以往大部分的公司都是採用固定維修時間及固定維修數量或經驗法則其中一種方式進行維修保養，某種特定比例關係來訂定維修保養的規則，依據曲線進行維修保養，設備在剛使用時，由於設備未經磨合，故障率很高；但隨著使用時間日益增加，故障率漸漸趨於穩定；在使用壽命終了時，故障率又逐漸增加，其故障率隨時間變化的關係如圖 1.3 所示，它的故障率兩頭較高，中間低，圖形就像浴盆一樣，所以稱為浴盆曲線。整理而言，業界認為若在耗損故障期發生之前就對設備進行檢驗，更換磨損的零件，就能防止其功能故障出現，但可能在固定維修保養時間來臨，已經發生當機現象，因而造成生產線，需要重新開機、維修調整等造成損失，影響整條

產線運作；生產線上的機台會因使用而磨損，需要適時更換，太早或太晚更換都可能造成損失，對企業而言額外的費用。因此，需更積極地做好保護措施，以避免企業承受重大損失。

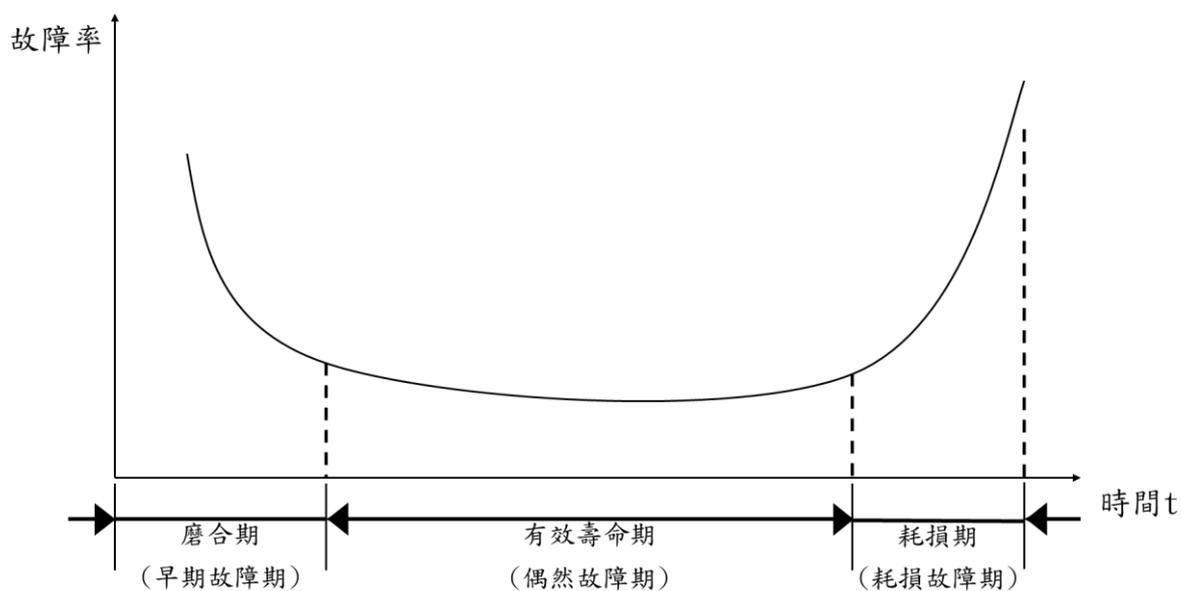


圖 1.3 浴盆曲線

(資料來源: Wikipedia)

## 1.2 研究動機

目前相關研究多半探討機台本身問題(例如:機台年齡)，缺少產能維修這方面研究。Stadje and Zuckerman (1996)提出設備會隨著時間而衰退，並降低生產力及收入，而保養可以使設備回到較佳的狀態，故需根據設備老化的狀態及保養程度發展最適合的維修保養策略。黃允成 (2017)提到設備使用後會有耗損或故障的現象，因此需進行預防維修保養，進而找出最適預防保養週期，降低成本影響。為了解決突發狀況當機、減少工廠之損失並因應需求變化，因此建構一符合現場當機狀況之預測模型有其必要性。

## 1.3 研究目的

為了解決突發狀況當機、減少工廠之損失並因應需求變化，因此建構一符合現場當機狀況之預測模型，為降低維護造成之設備可用時間損失又能兼顧產品之良率品質，本研究提出一三階段之「維護決策支援模組」，以資料挖礦手法結合 CPS 進行模擬，並透過實證案例驗證本研究之維護決策支援模組之有效性，提供企業做當機維修保養時程規劃參考依據。具體工作如下所述：

1. 利用分群法進行資料處理與資料檢視、
2. 透過決策樹進行維護規則之收斂、
3. 將規則投入 CPS 並模擬數據、
4. 比較規則生成前後之生產績效。

## 1.4 研究方法及步驟

本研究將應用於實際運作的晶圓製造廠，依照其製造廠實際運用之歷史資料進行資料探勘。本研究的流程依以下步驟進行，如圖 1.4 所示：

### 1. 定義問題及資料選擇

蒐集相關晶圓製造廠的製造流程，瞭解晶圓製造業維修保養之時程，利用資料庫中所需資料進行後續資料整理。

### 2. 資料前置處理

將所需資料進行前置處理，依據晶圓製造廠所提供資料及數據，來自不同地方的資料進行資料整合、資料清潔、資料轉換，資料處理完成後，進行資料分群處理，找出關鍵機群。

### 3. 決策樹建立

本研究針對前一步驟找出的關鍵機群做決策樹分類進行維修保養時程規劃預測。

### 4. 評估決策

依據決策樹執行結果歸納出法則，作為決策者在進行維修保養依據。

### 5. 模擬排程

依據決策樹執行結果，設定機台維修模組參數，並導入實際訂單需求，將法則設定至模擬排程中，機台當機與維修保養時程規劃，並給予維修建議。

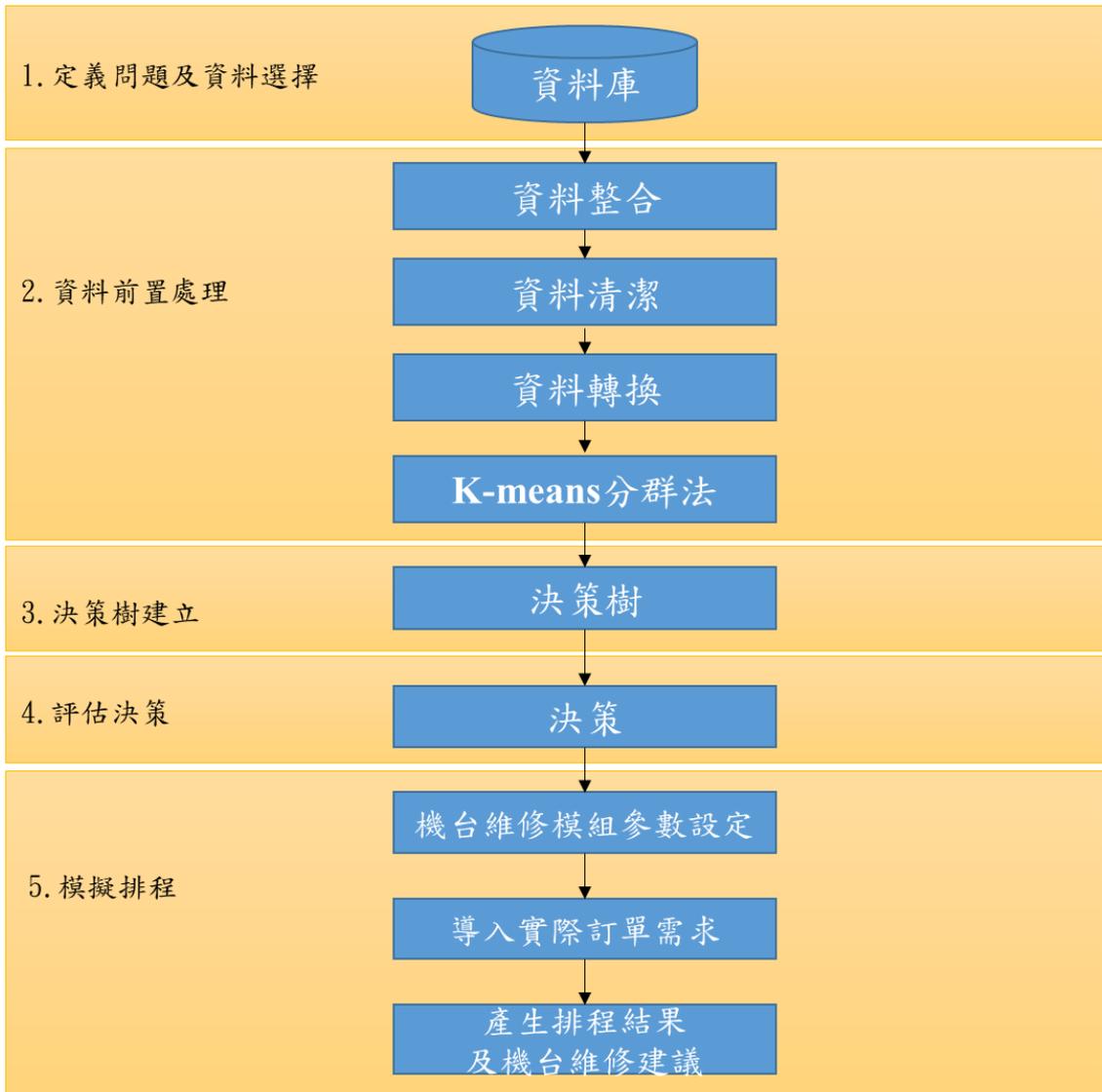


圖 1.4 本研究之架構與流程

## 第二章 文獻探討

### 2.1 維修保養策略

半導體建置時需大量的資金投入，設備是非常重要的資產，也是使得產品能夠有高附加價值的來源。機台因長時間使用運轉，會造成故障失效的情形，若能制訂一套完善的預防維護策略，在設備故障之前檢查、保養、修理該設備，或設備更新，可以使設備具有良好可用性，提高平均生產率，也能減少當設備不堪使用造成停機帶來的損失，以及無法如期交貨造成信譽受損與賠償等。就品質方面來說，發展適當的保養政策可保障工廠的生產效率，適度的維護與保養是不可或缺的，將損失降低為重要的一環，解決現今工廠所面臨的困難。

Lie and Chun (1986) 將維護分為兩類：修復性維修 (corrective maintenance, CM) 和預防性維護(preventive maintenance, PM)

一、預防維護(Preventive Maintenance, PM)：當設備在運作情況下，所進行之計畫性維護為減少當機可能性所採取的動作，皆稱為預防維護。包含兩種型式：

1. 簡單預防維護(Simple Preventive Maintenance)：設備於計畫時間或達到特定水準時對設備進行預防維護與保養動作。
2. 預防置換(Preventive Replacement)：設備經使用後老化至特定水準，對設備進行置換動作，以確保設備之正常運轉及性能之維持。

二、修正維護(Corrective Maintenance, CM)：屬於非計畫性行為，當設備於故障時才進行的修理動作，稱為修正維護。

1. 最小修復(Minimal Repair)：主要目的在消除設備已發生的故障，使其回復至故障前的狀態。
2. 修正置換(Corrective Replacement)：設備故障經修理後無法回復至可使用狀態，故進行置換更新的動作。
3. 完全修復(Complete (or perfect) Repair)：將產品的性能完全修復。

Moghaddam (2011)針對系統的預防保養策略及產品置換期間進行研究，假設產品可選擇更換、保養或維持現狀的方式且預防保養會受到產品所能提升的最大可靠度與成本預算的限制，對於每階段進行最佳化研究。Doostparast *et al.* (2014)考量可靠度下的預防保養計畫於連續性系統，使可靠度維修相關的成本最小化，成本函數包含：維修成本、更新成本、系統故障成本以及隨機失效故障成本。Wang and Tsai (2012)建立一個雙目標非完美 PM 模型的串並聯系統，其制定單位成本累積可靠性預期措施，以評估維持每個單獨部件在整個使用壽命期間的總體維護成本和系統可靠性的程度。Ebrahimipour *et al.* (2013)在多條生產線上開發了多目標 PM 調度模型，將生產線的可靠性以及系統的維護、故障和停機時間成本定義為多個目標，並對可用人力，備件庫存和維護期間採用不同的閾值。

Munirathinam and Ramadoss (2014)投入實際半導體資料(包括：Fault Detection and Classification (FDC)、Trace Log、Equipment Tracking (ET)、Metrology/Probe/Param 等資料)，使用的預測模型是神經網絡，決策樹或迴歸分析來得出結論，目的於通過模型運行傳入數據，並且實時生成資產健康分數，應用機器學習算法來生成數據驅動的資產模型。Munirathinam and Ramadoss (2016)建立一個決策模型並使用 WEKA 工具和 R 語言來實現提出的方法和五種機器學習技術，以幫助半導體業盡快檢測任何設備故障，以保持製造過程中的高生產率。良好的分類預測模型有利於半導體製造流程中的預測。大多數半導體製造非常複雜，且不斷製造出數百種測量數據，等待工程師進行分析，以保持高效運營並獲得高質量的最佳產量。對於如此大量的測量數據，數據分析技術至關重要。

## 2.2 資料探勘

資料探勘被定義為探索和分析資料的過程，透過自動或半自動的方式從大量資料中發現有意義的模式和規則(Witten *et al.*, 2016)。資料探勘也被許多人定義為從資料裡發現知識 (knowledge discovery from data, or KDD) 下圖為知識被發掘的步驟(Han *et al.*, 2011)包含七個步驟(如圖 2.1 所示):

- 1.資料清理(data cleaning)：移除雜訊。
- 2.資料整合(data integration)：將來自不同資料來源的資料整合。
- 3.資料選擇(data selection)：從資料庫中選取與目標相關之資料。
- 4.資料轉換(data transformation)：將資料轉換、編碼或簡化。
- 5.資料挖掘(data mining)：應用資料探勘技術萃取有用的知識。
- 6.評估(pattern evaluation)：利用衡量指標找出有用的資訊。
- 7.知識呈現(knowledge presentation)：利用視覺化的方式呈現。

因此，資料探勘是發現驅動的，不是由假設驅動。資料探勘應用於多個領域，用來處理大量的數據，包括企業資源管理(Jayaram and Singal, 2017)，教育(Romero and Ventura, 2013)，金融投資(Yan and Zheng, 2017)，健康和醫學領域(Chauhan and Jaiswal, 2016)，半導體製造業(Chien *et al.*, 2007)。資料探勘包括分類(classification)，找到將數據劃分為群組的規則；關聯(association)，找到在數據之間建立關聯的規則；順序(sequencing)，找到規則順序數據(Thuraisingham, 2014)。通常構建分類模型是從訓練數據中分析資料屬性和類別之間的關係，常被用來預測疾病或是行為的發生。關聯規則常被用來分辨資料庫裡屬性的相關性或相依性，這些規則是非常有用的已經在行銷、零售等多個領域被實證；找到關聯規則還能夠幫助我們發現一些更符合常理更正確的知識而這樣的發現有助於決策制定。順序模型，隨著時間序列找出事件同時發生的頻率，這些經常發生的事件模型是個人或計畫執行很重要的元素。

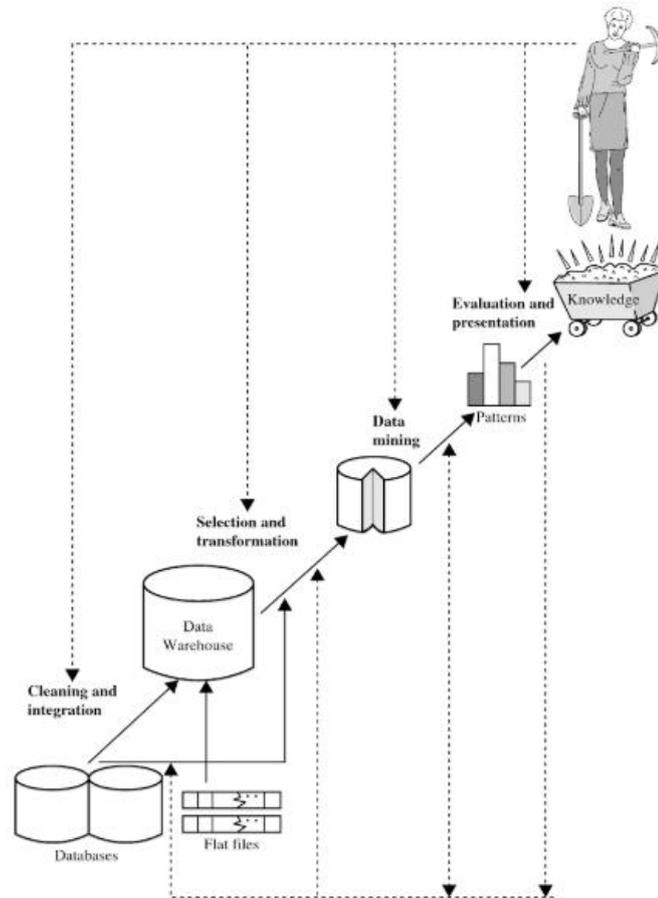


圖 2.1 資料採礦流程

資料來源：Han and Kamber (2011)

### 2.3 資料探勘應用於半導體

數據分析和機器學習方法在許多應用領域幾乎在每個行業中的加速增長，數據挖掘工具現在於半導體製造業中變得越來越流行。特別是在以收益為導向的增強技術中。這是因為傳統方法無法提取眾多複雜過程控制參數之間的隱藏關係，為了突出這些參數之間的相關性。Casali and Ernst (2012) 提出一個新的 KDD 模型，基於決策關聯規則和應急向量下，並運用 LHS-Chi2、post hoc 運算法，建立一個基於 KDD 模型的軟件，稱為 MineCor。此系統算法比較一般算法更加快速及精確，也可運用於晶體製造外的其他產業半導體技術持續升高，可靠度是去檢驗產品品質的關鍵步驟，由最終系統所收集的大數據可啟發數據挖掘的應用，用於新的可靠度數據分析方法。

Tyler (2016)提出一個新系統:RMIS,建立在工業 4.0 以及 IOT 方法下,並運用可靠度的提早預測,達成實驗室的全面訊息管理及大型可靠度測試的數據管理平台。依靠數據挖掘,收集可靠度數據,可進一步結合在線計量數據與失效分析去改進產能。掌握製造過程以及快速識別產量損失的原因是成功半導體製造工廠的關鍵。為理解產量損失原因,Barkia (2013)提出一個新的 KDD 方法:結合數據挖掘和整合工程師知識;有別於傳統的方法有兩個特點:

1. 分段並各自做分析,優點為通過分析各個分段可獲得第一次的分析,而連結各個關鍵部分就可以獲得更高階的分析。
2. 提出方法為多參數類聚分析,就像是基於多參數分析的群集一樣。

Yu (2016)提出在半導體產業中應用倒傳遞類神經網路(BPNN)來優化半導體晶圓的生產分配。利用數據挖掘的方法來預測工具集的到達率,並進一步開發BPNN模型來預測單工具集的WIP水準,最終達到規劃最小週期時間及優化生產資源,以下為兩種方法的解決方案。

1. 結合數據挖掘(Data mining)的空間統計和自適應共振理論,與缺陷相關聯來找到問題。
2. 倒傳遞類神經網路(BPNN)利用迴歸的神經網路(NN)來找到單個工具集及 WIP 水準之間的關係。

Capodici (2017)分析下列數據分析技術有無機器學習演算法之差別:

1. 傳統 DTCO 的主要缺點是只能考慮有限數量的評估和優化處理能力的佈局配置,換句話說,在製造 IC 產品的可預測性有限。
2. 物理設計覆蓋率(DSC)可以為實際的數據來分析計算框架。
3. 在網絡拓撲空間中可以在算法上識別共同點和根源。
4. 利用圖搜索和機器學習結合使用時可以將良品率做優化的方案。

本研究所探討的主題以半導體為主，近年使用資料採礦應用在半導體產業的研究整理如表 2.1 所示：

表 2.1 Data Mining 於半導體應用

研究者及發表年分	研究議題	研究方法
簡禎富、李培瑞、彭誠湧 (2003)	針對半導體製程製造測試的電性特徵萃取作為製程監控與事故診斷參考	決策樹、類神經網路
簡禎富、蕭禮明、王興仁 (2004)	半導體廠製造管理績效指標制定，協助使用者作為製造管理之依據	聚集、決策樹
Wang <i>et al.</i> (2005)	半導體封測廠中找出較佳的派工法則使整體完工時間達到最佳化	決策樹、類神經網路
Hsu & Chien (2007)	透過晶圓圖分類進行晶圓圖故障診斷	決策樹、分類規則
簡禎富，林昀萱，& 鄭仁傑 (2008)	半導體製程的良率	Kruskal-Wallis 檢定、卡方檢定、決策樹
Casali & Ernst (2012)	多個複雜的半導體製程控制參數間的相關性	卡方分析；決策樹
Chien, Hsu, & Chen (2013)	錯誤診斷與分類	多維主成分分析；K 最近鄰分析 (K nearest neighbor; KNN)
Wenjing, Yumin, Fei, & Xiang (2015)	半導體製造系統動態排程問題與分類	K 最近鄰分析 (K nearest neighbor; KNN)
Yu & Kuo (2016)	優化半導體晶圓製造中的生產資源配置	倒傳遞類神經網路 (BPN)
Hung, Lin, & Lee (2017)	半導體產品分層缺陷的問題	倒傳遞類神經網路 (BPN) 支持向量回歸 (SVR) 偏最小二乘法 (PLS)

(資料來源：本研究整理)

依據上面整理的表可以發現決策樹為資料探勘中常用工具之一，決策樹能夠找到顯著影響目標變數分布的屬性，透過選擇變數及指定目標將資料進行分類而成的樹狀架構，並能夠呈現出有層級架構之分類系統或是預測模型。

## 2.4 決策樹

決策樹是一種典型常用於分類和預測的數據挖掘方法，預測(prediction)技術依據某特定對象屬性，觀察歷史資料，用以推估其未來的可能性，應用在屬於監督式資料探勘，由於決策樹是將資料依據不同的變數循序來產生分析結果，可藉由決策樹分析方式來分析目標特質與異同點，提供決策者與其領域知識進行比較以驗證其決策並證明其合理性。決策樹具有易於解釋和理解的優點。除上述工具外，類神經網絡也能有分類的效果，決策樹該方法吸引人之處在於有規則的運用樹狀圖來表達資料規則的路徑，具有易於解釋和理解及能夠清楚展現研究目標及各項變數的關係之優點。

決策樹分析技術可同時用於連續型的資料數值(continuous variable)與類別型的資料數值(categorical variable)，並且還可適用於預測分類結果的分析(Tso and Yau, 2007；Jantan *et al.*, 2011)，表 2.2 為決策樹之比較。

C4.5 是由 Quinlan 於 1993 年提出，為 ID3 的升級版，克服資訊增益會選擇屬性時偏向選擇取值多的屬性的不足，改以增益比率做為選擇屬性的依據。

表 2.2 決策樹比較

倡者	Hartigan(1975)	Quinlan(1979)	Quinlan(1993)
演算法	CHAID	ID3	C4.5
資料屬性	類別型	類別型	類別型、連續型
節點分支	可有多分支	可有多分支	可有多分支
修剪規則	No pruning	Predicted error rate	Predicted error rate
分割規則	卡分檢定	Gain ratio	Gain ratio

決策樹利用監督式學習從訓練資料中，挑選出可以用以建構決策樹節點的屬性，決策樹包含決策節點、分支和葉節點三個部分，決策節點代表分類樣本的屬性，在該屬性上不同的節點代表一個分支，分支表示某決策節點的不同設定值，決策樹建構的步驟首先選擇一個屬性，利用屬性分類，計算分類後平均分散度，若分散度降得最低即為最佳分隔屬性，沒有屬性可以降低分散度或是只有一個數值時即停止。決策樹建立之後，可以用修剪方式 (tree-pruning) 步驟來縮減決策樹的大小，決策樹如果太大，很容易有過度學習的現象，所以修剪樹的動作將有助於改善決策樹類推至其他資料的能力。

決策樹優點如下點：

1. 決策樹適用於未處理過資料，相對其他方法較容易操作。
2. 決策樹分析結果能簡化為樹狀圖，較容易了解。
3. 可以處理連續性和類別性變項，不需太多的計算就可以進行分類。

#### 2.4.1 決策樹工具

Weka 由紐西蘭懷卡托大學基於 JAVA 環境下開源的機器學習以及資料探勘軟體，可以使用 java 語言開發更多資料採礦演算法，可以通過編程和命令列調用其分析組件。Weka 是一套完整的資料處理工具，包括對資料進行預處理、學習演算法、分類、迴歸及資料視覺化的使用介面，使用者可以快速的方法處理資料，同時還能比較和評估不同的學習演算法的性能，能夠找到解決問題最有效的方法。



圖 2.2 Weka 介面

(資料來源：Weka)

### 第三章 維修決策支援模組

本研究提出維修決策支援模組，利用企業提供之生產資訊，將資料進行預處理，資料整合、清潔、轉換，處理完成之資料，利用 K-means 技術來找尋關鍵機群。之後進行因子及參數設定後執行決策樹分析。因子設定(setting factor)為考慮使用者之現場維修保養之條件，分別為現場機台名稱、現場生產數量之累積、機台狀態(正常運作、當機及維修)、當機時間、維修保養時間及累積加工時間，針對這些因子去找尋適當的維修保養時程規劃；參數設定(setting parameter)用於協助資料預處理，幫助進入程式內可能需要做的格式轉換(定義資料型態)，調整信心因素值(confidence factor)、葉節點最小資料量(minNumObj)，調整這兩種參數能改善決策樹過度訓練(over fitting)的問題。將資料分成訓練組、測試組後決定資料輸出結果的設定。

決策樹根據葉節點目標變數的分布定義葉節點所代表的意義，呈現出具有層級架構之機群法則，一套適用於此機群的法則，此棵樹適用於此機群中各機台，依據決策樹所提出的法則歸納出的決策點可以給予現場工程師參考，並將法則參數設定至模擬系統中，判斷是否屬於法則區間中，如果是繼續排程到所有訂單完成為止即可結束模擬。

將決策樹所歸納出的法則存取至 SQL 中，並將模擬所紀錄之結果，其資料包含設備資訊、執行工單列表、日彙總表等，其記錄表單會由 ODBC 匯出外部資料庫。最後，比較紀錄表單當機比率的績效。

為兼顧生產效率與產品良率，本研究發展一個維修決策支援模組，主要分為三個階段，分別為規則生成階段、模擬結果階段與績效比較階段並以三個小節呈現說明。

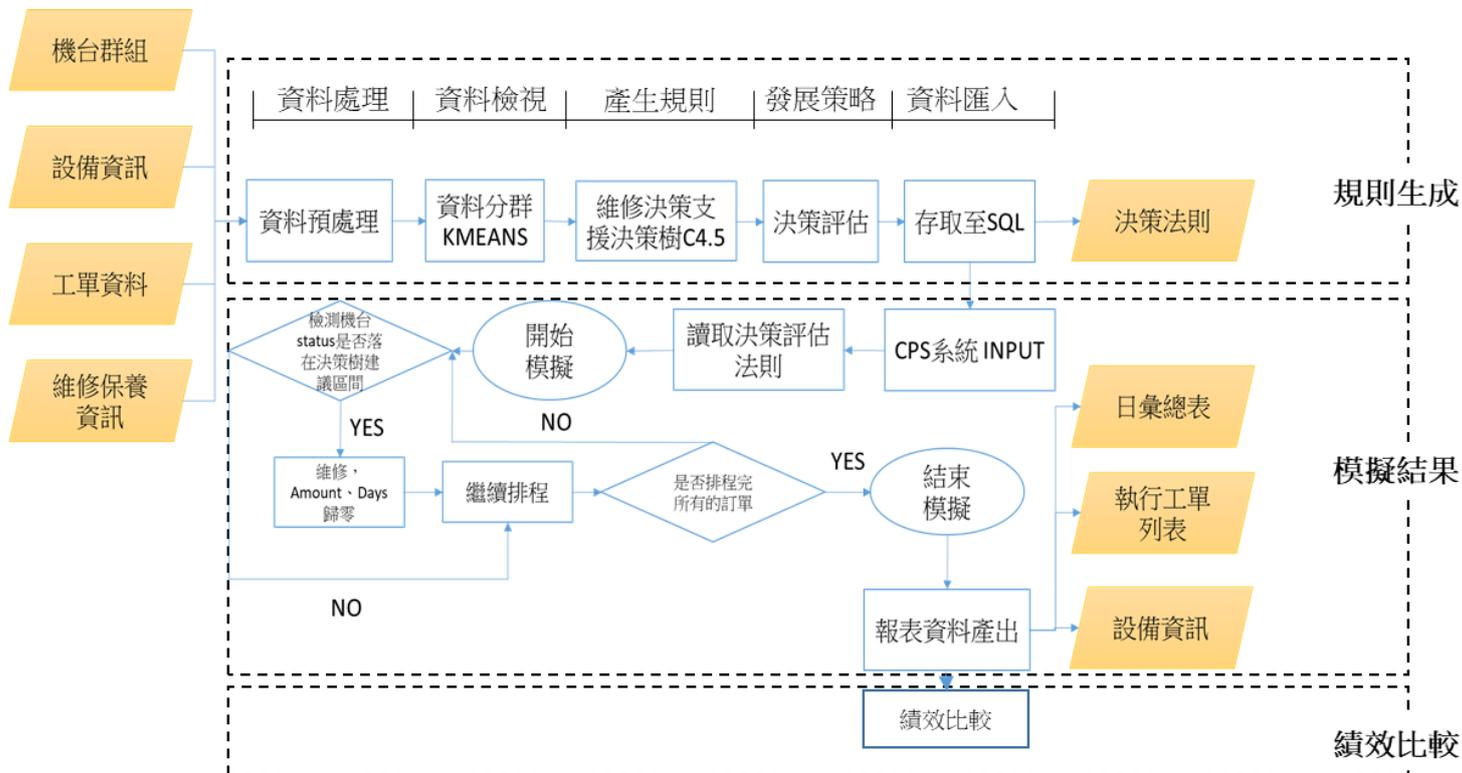


圖 3.1 維護決策支援模組流程圖

### 3.1 規則生成階段

本研究提出一維修決策支援模組之規則生成階段，主要分為五個步驟：資料處理、資料檢視、產生規則、發展策略與資料匯入。

#### 3.1.1 資料預處理

本研究應用於 M 公司十二吋晶圓廠，將透過現場歷史資料所產出之數據進行資料處理，通常取得的資料可能有許多變異、雜訊、缺失等問題，因此並無法直接使用，為了確保分析結果的品質及準確性，因此在進行資料探勘前須做資料的預處理，將來自不同資料源的資料進行合併，排除資料重複、遺漏的問題。本研究資料來自 MES、ERP 兩系統中，由 ERP 系統取得之物料相關資料(例如：訂單)；MES 得知機台設備相關資訊(例如：機台狀態、機台可用時間、維修資料)，將所蒐集之現場機台生產資訊與機台使用狀態資料，各機台資料及生產數量彙整，利用三個步驟，資料整合(data integration)、資料清潔(data cleaning)、資料轉換(data transformation)，進行資料預處理，以便後續分析。

本節將半導體晶圓廠的一機台進行案例說明。IHC06 的資料數據作為資料來源，根據廠商提供之資料來說明，

##### 1. 資料整合(Data Integration)

第一步驟資料前置處理，將來自不同資料庫資料進行資料整合，表 3.1 為企業提供之機台資料，EqpID 為機台名稱、E10 為機台狀態、Amount 為生產數量、StateName、ReasonCode 和 UserId 為企業內部紀錄的機台狀況名稱、原因代碼、操作人員，最後一個 StartTime 為狀態開始時間。

表 3.1 原始資料整合

EqpID	E10	Amount	StateName	ReasonCode	Userld	StartTime
IHC06	2	0	SBY-Standby	[AL0001] OK	14542	2013/01/01 17:20:32.455622
IHC06	1	5	PRD-Work for	C39206.0P from TCS		2013/01/01 19:48:09.066171
IHC06	2	0	SBY-Standby	[AL0001]	81374	2013/01/01 19:56:06.403357
IHC06	1	10	PRD-Work for	[AL0001]	81374	2013/01/01 19:56:19.606125
IHC06	3	0	TD experimen	[TS0008] down	00438	2013/01/01 20:00:19.606126
IHC06	2	0	SBY-Standby	C39206.0P from TCS		2013/01/02 09:58:45.262326
IHC06	1	21	PRD-Work for	C39331.05 from TCS		2013/01/02 10:30:01.404794
IHC06	2	0	SBY-Standby	C39331.05 from TCS		2013/01/02 10:45:24.845501
IHC06	1	84	PRD-Product	8A2438.00 from TCS		2013/01/02 12:59:42.023425
IHC06	2	0	SBY-Standby	8A2438.00 from TCS		2013/01/02 13:55:16.946394
IHC06	1	76	PRD-Product	8A2296.00 from TCS		2013/01/02 14:07:06.527962
IHC06	2	0	SBY-Standby	8A2214.00 from TCS		2013/01/02 15:05:35.571849
IHC06	1	35	PRD-Product	8A2359.00 from TCS		2013/01/02 15:51:58.223216
IHC06	2	0	SBY-Standby	8A2359.00 from TCS		2013/01/02 16:17:40.234381
IHC06	3	0	TD experimen	[TS0008] down	00438	2013/01/02 17:30:40.234381
IHC06	1	63	PRD-Engineer	840301.02 from TCS		2013/01/03 10:30:14.310818
IHC06	2	0	SBY-Standby	840301.02 from TCS		2013/01/03 14:21:57.876951
IHC06	1	22	PRD-Engineer	840301.02 from TCS		2013/01/03 15:23:17.869078
IHC06	2	0	SBY-Standby	840301.02 from TCS		2013/01/03 15:36:59.324442
IHC06	2	0	Wait For VIP	[AL0001] WAIT	13489	2013/01/03 15:39:24.359026
IHC06	1	54	PRD-Engineer	8A2245.00 from TCS		2013/01/03 16:49:26.544663
IHC06	4	0	SBY-Standby	840259.09 from TCS		2013/01/03 17:16:02.095130
IHC06	2	0	SBY-Standby	[AL0001] PM1	13489	2013/01/04 08:23:52.859161
IHC06	1	125	PRD-Work for	[AL0001] RUN	13489	2013/01/04 08:24:13.108702
IHC06	2	0	SBY-Standby	8A2342.00 from TCS		2013/01/04 11:02:40.042381
IHC06	1	61	PRD-Product	8A2511.00 from TCS		2013/01/04 11:03:38.320728
IHC06	2	0	SBY-Standby	8A2344.00 from TCS		2013/01/04 12:38:17.790691
IHC06	5	0	UDT-Maintenz	[DF0001] PM1 F	13707	2013/01/04 12:41:14.209947
IHC06	5	0	UDT-Equipme	[DF0005] PA:47	14657	2013/01/04 13:31:56.165188
IHC06	5	0	UDT-Normal q	[DF0001] PA:47	14657	2013/01/04 13:32:09.617102
IHC06	5	0	UDT-Wait for	[DF0001] PA:5e	14657	2013/01/04 13:50:44.753955
IHC06	2	0	SBY-Standby	[AL0001] PA:5e	14657	2013/01/04 13:50:59.100420
IHC06	1	23	PRD-Product	8A2541.00 from TCS		2013/01/04 14:00:33.479844
IHC06	2	0	SBY-Standby	8A2541.00 from TCS		2013/01/04 14:31:58.343061
IHC06	1	18	PRD-Engineer	840259.07 from TCS		2013/01/04 14:35:54.171908
IHC06	2	0	SBY-Standby	840259.07 from TCS		2013/01/04 14:54:24.605406
IHC06	1	1400	PRD-Engineer	840259.07 from TCS		2013/01/04 15:00:17.548970
IHC06	2	0	SBY-Standby	840259.07 from TCS		2013/01/04 15:10:20.538006

將來自不同資料庫資料合併後資料包含: EqpID、E10、Amount、StateName、ReasonCode、Userld 及 StartTime 共 7 項欄位，欄位說明如表 3.2 所示。依據表 3.1 舉例說明其意義，例如: 機台 IHC06 在 2013/01/01 17:20:32 455622 時，機台狀態為 2 等待，直到 2013/01/01 19:48:09 066171 時，開始正常運作，機台狀態為 1，而生產數量為 5。

表 3.2 欄位說明

欄位名稱	欄位說明	資料範例
<b>EqpID</b>	機台名稱	IHC06
<b>E10</b>	機台生產狀態	1 運作中 2 等待 3 關機 4 維修 5 當機
<b>Amount</b>	生產數量	10
<b>StateName</b>	企業內部紀錄的機台狀況名稱	SBY-Standby Time
<b>ReasonCode</b>	原因代碼	[AL0001]
<b>Userld</b>	操作人員	TCS
<b>StartTime</b>	狀態開始時間	2013/01/01 15:18:45

## 2. 資料清潔(Data Cleaning)

整合完成後，進行資料清潔，主要的工作包含：填補遺失值、處理遺失值問題，以及修正資料的不一致。面對遺失值，當蒐集的資料量很多，而遺缺資料為少數時，其影響性較小，便可直接刪除不進行分析。進行分類時，若目標屬性的分類標記為空白值，這筆資料因無法被正確分類，便可直接刪除。將原始資料表中的 StartTime 狀態開始時間的格式整理成最小單位為秒的形式，並減少資料集的資料量，刪除與本研究無關的屬性 StateName、ReasonCode 和 Userld 進行刪除，方便日後進行分析。本研究接續上述範例，從原始資料中取出需要的資料如機台名稱、機台狀態、生產數量以及狀態開始時間，並將 StartTime 狀態開始時間的格式整理成最小單位為秒的形式初步整理成為第二張資料表，如表 3.3：

表 3.3 資料清潔

EqplD	E10	Amount	StartTime
IHC06	2	0	2013/01/01 17:20:32
IHC06	1	5	2013/01/01 19:48:09
IHC06	2	0	2013/01/01 19:56:06
IHC06	1	10	2013/01/01 19:56:19
IHC06	3	0	2013/01/01 20:00:19
IHC06	2	0	2013/01/02 09:58:45
IHC06	1	21	2013/01/02 10:30:01
IHC06	2	0	2013/01/02 10:45:24
IHC06	1	84	2013/01/02 12:59:42
IHC06	2	0	2013/01/02 13:55:16
IHC06	1	76	2013/01/02 14:07:06
IHC06	2	0	2013/01/02 15:05:35
IHC06	1	35	2013/01/02 15:51:58
IHC06	2	0	2013/01/02 16:17:40
IHC06	3	0	2013/01/02 17:30:40
IHC06	1	63	2013/01/03 10:30:14
IHC06	2	0	2013/01/03 14:21:57
IHC06	1	22	2013/01/03 15:23:17
IHC06	2	0	2013/01/03 15:36:59
IHC06	2	0	2013/01/03 15:39:24
IHC06	1	54	2013/01/03 16:49:26
IHC06	4	0	2013/01/03 17:16:02
IHC06	2	0	2013/01/04 08:23:52
IHC06	1	125	2013/01/04 08:24:13
IHC06	2	0	2013/01/04 11:02:40
IHC06	1	61	2013/01/04 11:03:38
IHC06	2	0	2013/01/04 12:38:17
IHC06	5	0	2013/01/04 12:41:14
IHC06	5	0	2013/01/04 13:31:56
IHC06	5	0	2013/01/04 13:32:09
IHC06	5	0	2013/01/04 13:50:44
IHC06	2	0	2013/01/04 13:50:59
			•
			•
			•

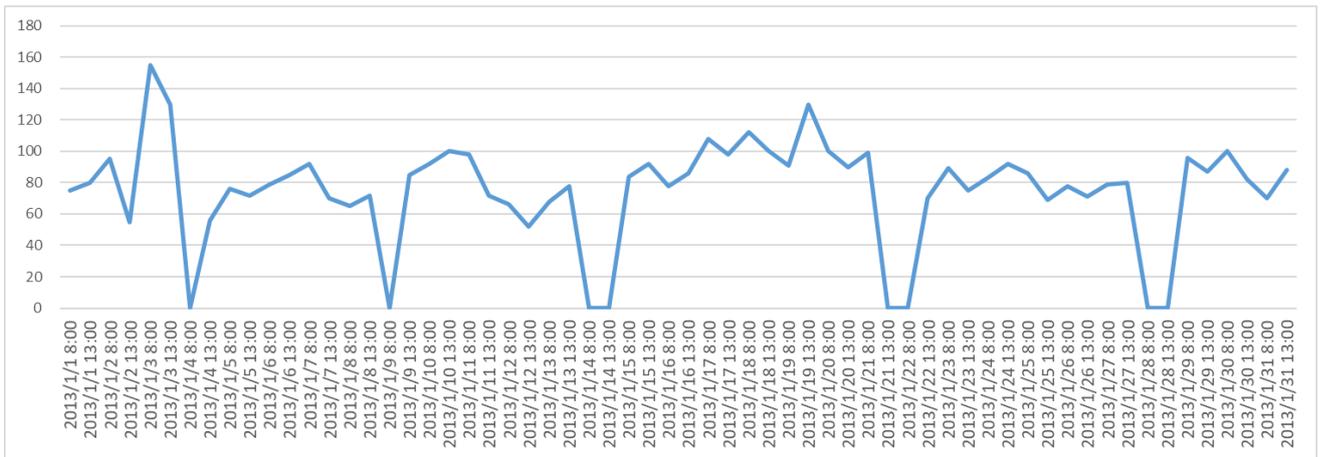


圖 3.2 時間-數量關係圖

根據表 3.3 繪製出圖 3.2 的時間-數量關係圖，可以觀察到除了 1/14 以及 1/28 的固定維修保養時間以外，在 1/4、1/9 發生了半天的當機以及 1/21 的全天當機，總共三次的突發當機。

### 3. 資料轉換(Data Transformation)

資料轉換是為了讓資料的數值在分析時不容易產生誤判，重點在資料數值要如何轉換，由於所蒐集之資料的值可能存在著單位格式不相同的差異，此會影響後續資料進行分析時的結果。依據不同資料探勘技術，將原始資料轉換成所需的資料格式。

本研究將原始資料表中，其研究變項之名稱、定義做了轉換如下，如表 3.4：

1. E10 的機台狀態列做轉換，將狀態 1、2 都分為正常運作，其餘以當機細分。
2. 將 Amount 做相加變成累積數量。
3. StartTime 原本為每個狀態的開始時間，將此時間轉換成累積加工時間天數，並將累積加工時間天數做區間轉換。

表 3.4 轉換欄位說明

欄位名稱	欄位說明
EqpID	機台名稱
Status	機台狀態分為： N正常運作 S短期當機(dt < 12小時) M中期當機(12 < dt < 24小時) L長期當機(dt > 24小時)
Amount	累積生產數量
Days	累積加工天數 1-3 天 → 1 4-6 天 → 2 7-10 天 → 3 11-14天 → 4

根據表企業提供之資料可得機台維修保養的資料，可得知目前的機台為兩週進行一次的維修保養，但在兩週的時間內，機台可能因種種因素造成機台當機。因此本研究利用決策樹找出在兩次定期維修之間發生的所有當機事件之規則。將資料先進行數轉換，並將投入因子時間以天數的格式累積，累積至機台定期維修保養以及不定時的小維修保養(10分鐘內可以完成的動作，對產能影響極小的)。即為兩種情境：

1. 當天數累積至 14 天時就會歸零，因為每兩週有一次定期保養
2. 當看到表格機台狀態列出現狀態 4 時，也會停止累積

而第二個投入因子也是照著同樣的規則來整理，依照時間的累積同時累積在這段時間內生產的產品數量。依照上述的規則整理表 3.4 後可以得到表 3.5。

表 3.5 資料轉換

Amount	Day(s)	status
305	1	N
320	1	S
382	2	S
400	1	M
330	3	N
327	1	S
328	2	S
330	1	N
331	2	N
335	2	M
335	1	M
337	2	M
338	1	S
340	1	N
340	2	S
340	2	S
341	2	S
341	1	S
342	1	S
342	3	N
	.	
	.	
	.	

表 3.5 的 Amount 為累積的生產數量、Day(s)為累積的天數、status 為機台狀態。以表 3.5 第一列為例當累積加工 305 個產品並且是用 1 天的時間加工完成後並沒有當機，所以機台狀態為 N，而停止累積的原因為遇到不定期的小保養。

### 3.1.2 資料檢視

本研究應用分群(Clustering)技術進行資料檢視。分群為將資料分類成群的一種方法，可以找出某些隱藏於資料內未知的關係，找出其所存在的規律性，將較相似的資料分群在一起，做為進一步預測與決策時的參考依據。分群分析法通常可分為分層法 (Hierarchical)、非分層法 (Nonhierarchical)和兩階段法。

而本研究應用 K 組平均法(K-Means)並配合專家判斷找尋出實驗對象的“關鍵”機台群組，針對關鍵機群進行後續分析找出其法則。

首先，將整理好之過去歷史生產資料，針對特定指標(例如:loading)對產能影響進行分析，將所有對象分為 K 個簇(群)，以使簇(群)內的相似度較高，分群找出關鍵機台。

步驟 1：將整理好之過去歷史生產資料，並進行資料前處理，設定 K-means 分群數量。

步驟 2：執行 K-means 演算法，所有對象分為 K 個簇(群)。

步驟 3：以使簇(群)內的相似度較高，分群找出關鍵機台，實驗所需進行機台數判定所需群數，找到可能為“關鍵”機台群組進行後續實驗。如圖 3.3，為 K=3 的 clusters。

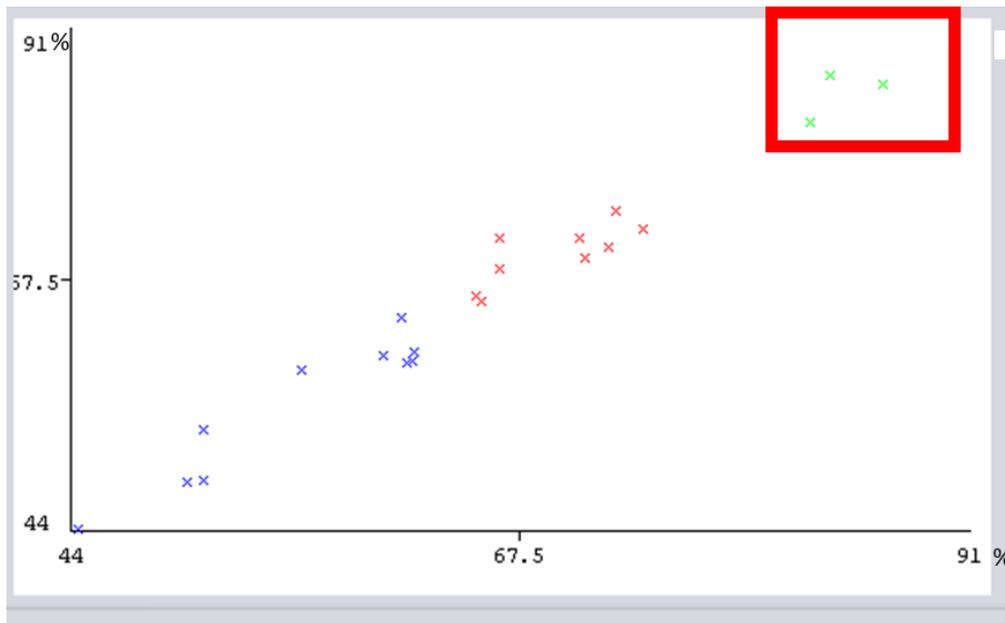


圖 3.3 分群法

以圖 3.3 為例，紅色框內的機台群組，成為“關鍵”機台群組的可能性較大，找到可能為“關鍵”機台群組進行後續分析。

本小節定義出可能影響產能之相關指標，並進行初步的資料整理以配合後續案例分析使用；同時利用群聚分析方法篩選出所需使用之“關鍵”機台群組在進而使用專家法，看分完這個紅色中是否有測試機跟實驗機，如果分群結果中有此兩類機台，將此兩類機台進行移除。

### 3.1.3 決策樹發展規則

部分學者認為決策樹所發展法則的能力較好，適合用於找出法則協助診斷，因為決策樹得到的結論能有很好的解釋能力，比較容易找出投入項目與維修保養之間的關聯規則，較適合本研究。決策樹說明如下所示：

1. 資訊理論為假設有一資料庫  $D$  共有  $|D|$  筆資料，這  $|D|$  筆資料可以分類為  $k$  種結果，這  $k$  種結果發生的機率分別為  $p_1, \dots, p_k$ ，這些機率都是已知，則這個事件發生後所得到的資訊量：

$$Info(D) = \sum_{i=1}^k -p_i \log_2 p_i \quad (1)$$

2. 資訊增益(Information Gain):

$$Gain(A) = Info(D) - Info_A(D) \quad (2)$$

在決策樹分類的演算法，如 CART、CHAID、ID3、C4.5，本研究將採用 C4.5 演算法進行資料分析，Quinlan 於 1993 年以 ID3 演算法為基礎，改良出新的決策樹演算法，稱為 C4.5 演算法。其演算法基於資訊理論 (Information Theory) 的方法，以資訊理論為基礎，利用資訊增益率 (Gain Ratio) 來選擇屬性，資訊增益率(Gain Ratio)來取代原有的分岔準則，從而實現對數據的歸納分類，以下為資料切割之依據：

1. 考量資料集合切割時所產生的切割資訊量(split information)，資料集 D 中，屬性集合  $A = \{ ID, Days, Amount \}$ ，通過將資料集 D 劃分成對應於屬性 A 測試的  $v$  個劃分產生的資訊， $j$  為劃分標號，其定義如下：

$$SplitInfo_A(D) = -\sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \frac{|D_j|}{|D|} \quad (3)$$

2. 資訊增益率(Gain Ratio)的公式定義如下：

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo_A(D)} \quad (4)$$

C4.5 演算法流程如圖 3.4 所示，在每一次創建節點 N 時，此時所有物件都屬於 N 的物件集合。若 N 中所有的物件都屬於同一類別，則定義 C 為此類別並停止。對屬於 C 的所有物件，分別計算其切割資訊量，從根節點至目前節點中，並分別計算部分決策樹的切割資訊量及資訊增益率。資訊增益率最大的屬性即為當前節點的分裂屬性，隨著遞迴計算。選擇具有最大資訊增益率的屬性，並當做 N 節點的分類屬性。最後利用後剪枝方法進行剪枝，決策樹完成。

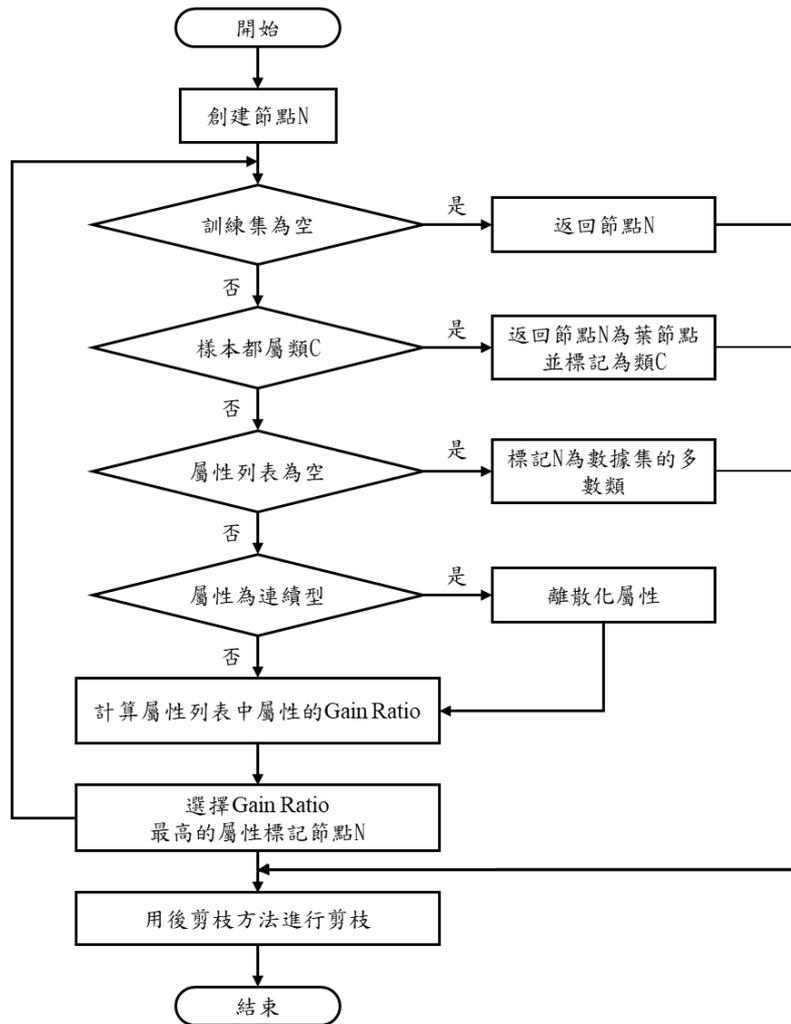


圖 3.4 C4.5 流程

當決策樹分支數目過多時，可能導致過度擬合問題，會使得準確度降低，因此需進行決策樹修剪，減少無效枝的數量，而在建構決策樹時，其停止成長的條件為：

1. 節點內的數據已屬於同一類別
2. 節點內的數據樣本數小於可分割值
3. 所有屬性已被分裂過

C4.5 根據決策樹的每個葉子節點生成一個法則集，對於法則集中的每條規則，從樹底部往上來嘗試是否有條件可以移除，採用後剪枝方式以錯誤率來作為判斷剪枝之依據，當訓練資料中有  $N$  個樣本，當中有  $E$  筆錯

誤訓練資料分類錯誤，則其預測錯誤率則可以用  $(E+0.5)/N$  表示，0.5 為一個調整係數；

對於一棵擁有  $L$  個葉子的子樹，則子樹的錯誤率為  $e$ ：

$$e = \frac{(\sum E + 0.5 \times L)}{\sum N} \quad (5)$$

當子樹以葉節點代替後所得到的錯誤率較低則修剪掉該子樹，並以葉節點代，如果錯誤率比較高則保留該子樹。

依據流程建立出決策樹，從中學習其法則，每個內部節點表示在一個屬性上的測試，每個樹葉節點存放一個類標籤，一旦建立好了決策樹，對於一個未給定類標籤的類別，走一條有根節點到葉節點的路徑，該葉節點就存放著該類別的預測。當我們在建立分類模型時，除了預測的準確度要高外，也希望能了解其分類的依據，並以簡單易懂的方式所呈現。

本研究提供一案例說明。依照上述流程進行資料處理完成後，這邊使用機台資料集中的 50 筆當機資料(正常 10 筆；當機 S17；當機 M7 筆；當機 L 16 筆)來舉例：

計算 Gain Ratio 根據以公式(6)來計算母節點的資訊量：

$$Info = -\frac{10}{50} \log_2 \frac{10}{50} - \frac{17}{50} \log_2 \frac{17}{50} - \frac{7}{50} \log_2 \frac{7}{50} - \frac{16}{50} \log_2 \frac{16}{50} = 1.9167 \quad (6)$$

若使用「Amount」來做為分岔變數。利用這個變數可以將母節點切成兩部分。第一部分是「 $\leq 375$ 」，共計 19 個，有當機有 4 筆；第二部分是「 $> 375$ 」，共計 31 個，有當機有 11 筆。因此，使用「Amount」來做為分岔變數呈現如公式(7)：

$$Info = \frac{19}{50} \left( -\frac{4}{19} \log_2 \frac{4}{19} - \frac{15}{19} \log_2 \frac{15}{19} \right) + \frac{31}{50} \left( -\frac{20}{31} \log_2 \frac{20}{31} - \frac{11}{31} \log_2 \frac{11}{31} \right) \\ = 0.8212 \quad (7)$$

則 Gain 計算如公式(8)：

$$Gain = 0.9988 - 0.8212 = 0.1766 \quad (8)$$

在這個案例中，透過「Amount」將資料由 50 筆切割 19 筆、31 筆兩部分，切割資訊量(split information)計算如公式(9)：

$$SplitInfo = -\frac{19}{50} \log_2 \frac{19}{50} - \frac{31}{50} \log_2 \frac{31}{50} = 0.958 \quad (9)$$

則資訊增益率計算如公式(10)：

$$Gain Ratio = \frac{0.1766}{0.958} = 0.1409 \quad (10)$$

依據上述步驟反覆計算後，得到所有屬性的增益率(Gain Ratio)，將增益率由高到低排序，增益率最高者適合做為決策樹第一層的分岔變數，可以得到一棵完整的機台當機時程規劃預測決策 (見圖 3.5 的決策樹)。

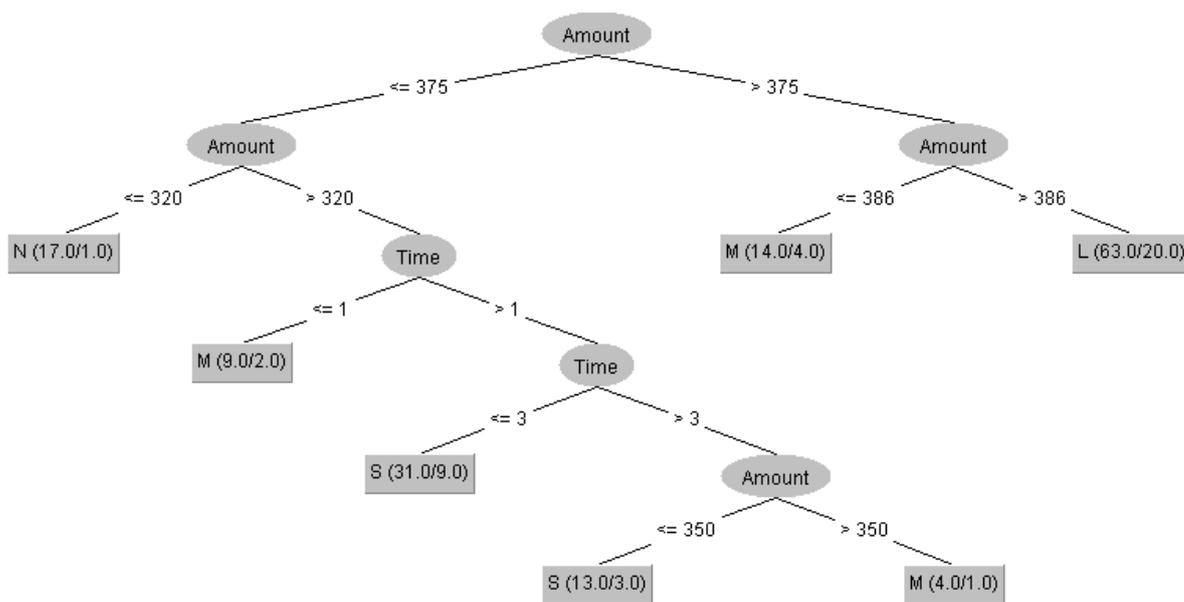


圖 3.5 範例決策樹

依照圖 3.5 可以歸納出如表 3.6 所示的決策樹法則：

表 3.6 模式範例法則

編號	規則	說明
1	Amount > 386	當生產數量大於 386 時，為長期當機
2	386=>Amount > 375	當生產數量介於 386 到 375 間，為中期當機
3	375=> Amount >350； Time >3	當生產數量介於 375 到 350 間，且加工天數大於 3 天時，為中期當機
4	Amount < 350； Time >3	當生產數量小於 350，且加工天數大於 3 天，為短期當機
5	375=> Amount > 320； Time <=1	當生產數量介於 375 到 320 間，且加工天數小於 1 天時，為中期當機
6	Amount <=320	當生產數量小於 320 時，為正常運作

將資料分成兩個資料集為訓練資料及驗證資料，並使用訓練資料進行模式的開發，再透過驗證資料去進行模式的評估，最後預測準確度，將預處理後的資料以隨機抽樣的方式以 80:20 的比例，分為訓練資料及驗證資料，可以發現此決策樹正確率為 73.3%，如圖 3.6 所示。

=== Summary ===

Correctly Classified Instances	22	73.3333 %
Incorrectly Classified Instances	8	26.6667 %

圖 3.6 決策樹正確率

經由上述之維修策略法則，本研究將其整理成一維修策略表格，如表 3.7 所示；此表格將會整理出所有機台的維修策略，此表格將會整理出所有機台的維修策略，機台名稱(Machine Name)能夠辨別出不同機台的策略，累積數量的區間(Amount Upper/Lower Bound)表示機台應當發生維修的數量區間，而累積加工時間區間(Time Upper/Lower Bound)表示機台應當發生維修的累積時間區間，MTTR(Mean-Time-to-Repaired)表示維修保養所需時間，而維修當機的機率是利用決策樹的決策數量以及反決策數而計算出的機率，決策數量總數與反決策數分別記錄於 CONT 與 N-CONT。發生機率

(Probability)為使用者可自行設定之機率，若上述維修當機的機率未超過發生機率(Probability)所設定之機率，則不列入維修時程規劃中。

我們此表單第五欄作為範例，以 IHC01 的資料數據作為資料來源，Amount Upper Bound 為 375，Amount Lower Bound 為 320，表示此機台應在數量 320-375 數量之間即進行維修；Time Upper Bound 為 1:00:00:00，Time Lower Bound 為 0:00:00:00，表示此機台應在時間 0-1 天之間即進行維修。我們可得知當生產數量介於 375 到 320 間，且加工天數小於 1 天時，為中期當機，因此 downtime 顯示為 M。並且當 MTTR 為維修 20 小時、機率為 0.77，其決策數量總數與反決策數分別 9 與 2。

因此可以透過各機台的累積數量與累積加工時間得出其當機的模式與可能機率為何，並得知決策數量總數與反決策數的數量。

表 3.7 決策樹維修策略表

Machine Name	Amount Upper_Bound	Amount Lower_Bound	Time Upper_Bound	Time Lower_Bound	MTTR(hr)	Probability	DownType	CONT	N-cont
IHC01		386			115	0.682539683	L	63	20
IHC01	386	375			15	0.714285714	M	14	4
IHC01	375	350		3:00:00:00	22	0.75	M	4	1
IHC01	350			3:00:00:00	5	0.769230769	S	13	3
IHC01	375	320	1:00:00:00	0:00:00:00	20	0.777777778	M	9	2
IHC01	375	320	3:00:00:00	0:00:00:00	3	0.709677419	S	31	9
IHC01	320				0	0.941176471	N	17	1

## 3.2 模擬結果階段

### 3.2.1 CPS 系統架構

機台維修保養是一門學問，如果能適當的排出維修保養時程規劃，就能減少產能及成本的損失，因此本研究考慮工廠實際的維修保養狀況及生產情形之維護決策支援模組架構，為了解決工廠當機情況、減少工廠之浪費，並考慮工廠實際生產問題。本文提出的維護決策支援模組架構是基於 Wang *et al.*(2016)的 Capacity Planning Simulation (CPS)系統進行發展，

CPS 系統是由五個主要模組所組成：(1) 資料輸入、(2) 模擬系統、(3) 產能規劃、(4) 維護決策支援系統和 (5) 資料輸出。每個模組的作用和功能描述如下。

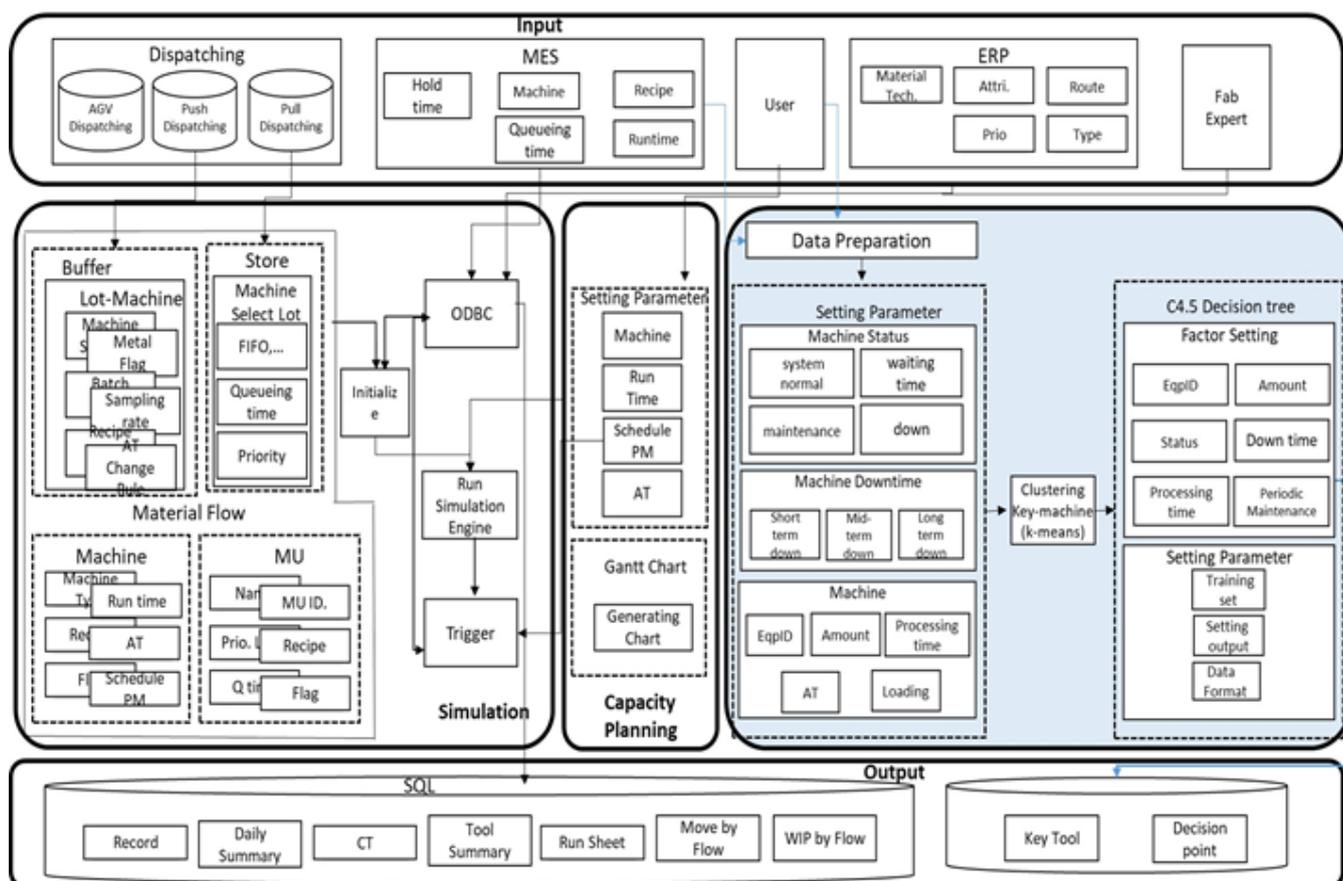


圖 3.7 CPS 系統架構

(資料來源: Wang *et al.*, 2016)

## 1. 資料輸入

資料輸入模組包含資料和邏輯，其資料由 ERP 系統所得之物料相關資訊(例如:訂單資訊、產品種類)；由 MES 系統提供機台相關的數據(例如:機台狀態、機台運行時間)；邏輯則有工廠使用生產的限制(例如: Q-Time) 及派工法則。

## 2. 模擬模組

模擬模組由模擬引擎、模擬法則控制、模擬參數控制所組成。根據產品訂單、半導體晶圓廠的生產特性、生產限制、派工法則及參數設定；參數分兩類，物料參數其內容包含物料的種類(Tech、Route)、優先級別(Priority)、屬性(Flag、Q-Time、Sampling Rate)；而機台參數則為可調參數包括 Recipe、機台種類(Machine)、機台加工時間(Runtime)、可用時間(AT)。模擬模組能準確模擬每個批次/產品的製造過程及每個機台的生產狀況。

## 3. 產能規劃

產能規劃模組主要為用戶提供一個介面來規劃和控制參數(例如: AT、維修保養規劃)。將啟動模擬模組來產出生產/操作時間表，其通常充分利用可用產能來滿足每週/每月需求計劃(例如: 晶片產出量)。

## 4. 維護決策支援

維護決策支援模組將資料進行資料預處理，根據產品訂單、機台設備相關資訊(例如:機台執行時間、機台狀態、機台可用時間及維修保養時程資訊)，找尋出關鍵機群，針對關鍵機群進行因子設定，選定目標因子後執行決策樹，利用決策樹歸納出一套法則，提供現場作為維修保養時程規劃之依據，配合模擬系統，歸納出來之法則列入排程中，調整原先的維修保養時程規劃。

## 5. 資料輸出

輸出之結果分為模擬系統所產出之結果資訊包含 run sheet、週期時間、最晚投料時間、最晚完成時間等，以表 3.8 說明模擬產出內容，記錄每一個訂單生產流程資訊，了解物件製造程序，包含製造何種製程，在其機台運行時的開始加工時間、加工之數量、加工結束時間及其週期時間，以此檢測生產模擬情況。另一種則是維護決策支援結果(例如:關鍵機群、決策點(法

則)，通過 ODBC 匯出至外部資料庫，並儲存於 SQL 當中，可將檔案匯出並利用 Excel 表示結果。

表 3.8 品項生產資訊表單

LotID	PartID	OPENum	QueueTime	TrackIn	TrackOut	Priority	SOD	Machine	CR	QT	RunTime	CycleTime
CFG95612.1	PD 1	680.000	2017-10-04T12:39:45Z	2017-10-04T12:39:45Z	2017-10-04T13:52:30Z	25	2017-10-05T06:00:00Z	PAG 1	-2.727772829074621	1970-01-01T00:00:00Z	00:01:12:45	00:01:12:45
CFG95612.1	PD 1	670.000	2017-10-03T22:10:00Z	2017-10-04T12:34:45Z	2017-10-04T12:39:45Z	25	2017-10-05T06:00:00Z	MAC 1	-2.932884429400387	1970-01-01T00:00:00Z	00:00:05:00	00:00:05:00
CFG95612.1	PD 1	660.000	2017-10-03T22:00:00Z	2017-10-03T22:00:00Z	2017-10-03T22:10:00Z	25	2017-10-05T06:00:00Z	IQC 1	-5.385060572126642	1970-01-01T00:00:00Z	00:00:10:00	00:00:10:00
CFG95612.1	PD 1	650.000	2017-10-03T16:50:00Z	2017-10-03T21:53:30Z	2017-10-03T22:00:00Z	25	2017-10-05T06:00:00Z	SOR 2	-5.4132546065356815	1970-01-01T00:00:00Z	00:00:06:30	00:00:06:30
CFG95612.1	PD 1	640.000	2017-10-03T16:45:00Z	2017-10-03T16:45:00Z	2017-10-03T16:50:00Z	25	2017-10-05T06:00:00Z	MAC 1	-6.287269673215922	1970-01-01T00:00:00Z	00:00:05:00	00:00:05:00
CFG95612.1	PD 1	630.000	2017-10-03T16:15:00Z	2017-10-03T16:15:00Z	2017-10-03T16:45:00Z	25	2017-10-05T06:00:00Z	MIC 6	-6.301366690420442	1970-01-01T00:00:00Z	00:00:30:00	00:00:30:00

維修決策支援模組輸入資料由 MES、ERP 系統所取得，以下舉例兩張表單分別說明投入資料。透過表 3.9 可以了解每機台來自哪一群組及其加工時間、維修所需時間。表 3.10 為客戶訂單，分別生產兩種產品 RSH\_1 及 RSH\_2，其產品之製程、生產數量、客戶交期。

表 3.9 機台群組

Tool Name	AT	MTR	Tool_Group	Sub_TG	Upperbound	Lowerbound
SOR_1			TGSOR			
SOR_2			TGSOR			
SOR_3			TGSOR			
STH_1			TGSTH			
STH_2			TGSTH			
SCRA_1			TGSCRA			
SCRA_2			TGSCRA			

表 3.10 訂單資訊

Customer	LOTID	Tech	Start Operation	End Operation	QTY	PRIORITY	CLAIM TIME	SOD
Cus_1	SD_1	RSH_1		010.00	3000	1	2018/5/29 08:30:00	2018/6/4 17:00:00
Cus_1	SD_1	RSH_2		010.00	3000	1	2018/5/29 08:30:00	2018/6/8 17:00:00
Cus_2	SD_2	RSH_1		010.00	3000	2	2018/5/29 08:30:00	2018/6/15 08:30:00
Cus_2	SD_3	RSH_2		010.00	3000	2	2018/5/29 08:30:00	2018/6/10 15:30:00
Cus_3	SD_4	RSH_1		010.00	3000	3	2018/5/29 08:30:00	2018/6/5 17:00:00

當執行完決策支援模組時，產出之法則存於 SQL 中。將決策支援模組產生之結果輸出，作為進行模擬排程前的資料輸入。即進行模擬排程時，再從 SQL 讀取法則，依據決策樹歸納出來的法則進行參數設定，填入表 3.7 中，將維修保養列入模擬排程中做規劃，調整原先規劃之維修保養時程。

本研究以範例執行上述流程，在確認決策樹的可行性後，從 ERP 內取得一張新的訂單，如表 3.11。此訂單生產兩種產品 RSH\_1 及 RSH\_2，將這兩項產品之製程、機台資訊等等輸入 CPS 模擬排程系統之輸入表單。

表 3.11 客戶訂單

Customer	LOTID	Tech	Start_Operation	End_Operation	QTY	PRIORITY	CLAIM_TIME	SOD
Cus_1	SO_1	RSH_1			500	1	2018/5/29 08:30:00	2018/6/4 17:00:00
Cus_1	SO_1	RSH_2			200	1	2018/5/29 08:30:00	2018/6/8 17:00:00
Cus_1	SO_2	RSH_1			100	2	2018/5/29 08:30:00	2018/6/15 08:30:00
Cus_2	SO_3	RSH_1			50	2	2018/6/2 13:15:00	2018/6/8 15:30:00
Cus_2	SO_3	RSH_2			100	2	2018/6/3 13:15:00	2018/6/10 15:30:00

CPS 模擬排程系統輸入表單包含表 3.12 產品製程(TECH 產品名稱、Oper Num 製程編號、ToolG 機台群組)、表 3.13 製程時間(Tool\_ID 機台名稱、Tech 產品名稱、OP No 製程編號、Runtime 製程時間)、表 3.14 機台群組(Tool Name 機台名稱、Tool\_Group 機台群組)以及表 3.15 機台開關機時間設定(Tool Name 機台名稱、Up/Down 機台循環開關機時間、AbsUp/AbsDown 機台維修保養時間)。

表 3.12 產品製程表

TECH	Oper.Num	ToolG
RSH_1	010.00	IHC
RSH_1	020.00	FG
RSH_1	030.00	FH
RSH_1	040.00	GA
RSH_1	050.00	FI
RSH_1	060.00	CM
RSH_1	070.00	HO
RSH_1	080.00	GO
RSH_1	090.00	AS
RSH_1	100.00	AS
RSH_1	110.00	PK
RSH_2	010.00	IHC
RSH_2	020.00	FG
RSH_2	030.00	FH
RSH_2	040.00	GA
RSH_2	050.00	FI
RSH_2	060.00	CM
RSH_2	070.00	HO
RSH_2	080.00	GO
RSH_2	090.00	AS
RSH_2	100.00	AS
RSH_2	110.00	PK

表 3.13 製程時間表

Tool_ID	Tech.	OP No	Runtime
AS_1	RSH_1	090.00	100
AS_1	RSH_1	100.00	120
AS_2	RSH_1	090.00	100
AS_2	RSH_1	100.00	130
CM_1	RSH_1	060.00	450
CM_2	RSH_1	060.00	490
CM_3	RSH_1	060.00	290
FG_1	RSH_1	020.00	860
FG_2	RSH_1	020.00	800
FH_1	RSH_1	030.00	870
FI_1	RSH_1	050.00	560
GA_1	RSH_1	040.00	550
GO_1	RSH_1	080.00	24000
HO_1	RSH_1	070.00	18000
PK_1	RSH_1	110.00	300
IHC_1	RSH_1	010.00	240
IHC_1	RSH_1	010.00	240
AS_1	RSH_2	090.00	100
AS_1	RSH_2	100.00	100
AS_2	RSH_2	090.00	100
AS_2	RSH_2	100.00	150
CM_1	RSH_2	060.00	450
CM_2	RSH_2	060.00	500
CM_3	RSH_2	060.00	300
FG_1	RSH_2	020.00	850

表 3.14 機台群組表

Tool Name	Tool Group
IHC_1	IHC
SC_1	SC
FG_1	FG
FG_2	FG
FH_1	FH
GA_1	GA
FI_1	FI
CM_1	CM
CM_2	CM
CM_3	CM
HO_1	HO
GO_1	GO
AS_1	AS
AS_2	AS
PK_1	PK

表 3.15 機台開關機時間設定表

Tool Name	Up	Down	AbsUp	AbsDown
IHC_1	8:00:00	17:30:00		
IHC_1			6/2/2018 17:30	6/2/2018 8:00
IHC_1			6/16/2018 17:30	6/16/2018 8:00
SA_2	8:00:00	17:30:00		
SC_1	8:00:00	17:30:00		
FG_1	8:00:00	17:30:00		
FG_2	8:00:00	17:30:00		
FH_1	8:00:00	17:30:00		
GA_1	8:00:00	17:30:00		
FI_1	8:00:00	17:30:00		
CM_1	8:00:00	17:30:00		
CM_2	8:00:00	11:30:00		
CM_3	0:30:00	7:30:00		
HO_1	8:00:00	17:30:00		
GO_1	8:00:00	17:30:00		
AS_1	8:00:00	17:30:00		
AS_2	6:30:00	11:30:00		
PK_1	8:00:00	17:30:00		

從表 3.15 機台開關機時間設定表設定了每兩週一次的固定維修保養時間，分別在 2018/6/2 8:00 至 2018/6/2 17:30 以及兩週後的同一個時段 2018/6/16 8:00 至 2018/6/16 17:30，將上述五張表投入 CPS 模擬系統內進行生產排程後可以得到以下的結果表 3.16 生產紀錄表(LotID 批號、PartID 產品編號、OPENum 製程編號、QueueTime 等待時間、TrackIn 進機時間、TrackOut 出機時間、Priority 優先等級、SOD 交期時間、Machine 機台名稱、Quantity 生產數量)、表 3.17 機台資訊紀錄表(Date 日期、ToolID 機台名稱、ToolGroup 機台群組、Run 機台運作時間比例、Idel 機台閒置時間比例、Down 機台當機時間比例、ChangeLine 機台換線時間比例、UTIL 機台稼動率)。

表 3.16 生產紀錄表

LotID	PartID	OPENum	QueueTime	TrackIn	TrackOut	Priority	SOD	Machine	Quantity
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T08:30:00Z	2018-05-29T09:10:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T08:45:00Z	2018-05-29T09:25:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T09:10:00Z	2018-05-29T09:50:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T09:25:00Z	2018-05-29T10:05:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T09:50:00Z	2018-05-29T10:30:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T10:05:00Z	2018-05-29T10:45:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T10:30:00Z	2018-05-29T11:10:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T10:45:00Z	2018-05-29T11:25:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T11:10:00Z	2018-05-29T11:50:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T11:25:00Z	2018-05-29T12:05:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T11:50:00Z	2018-05-29T12:30:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T12:05:00Z	2018-05-29T12:45:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T12:30:00Z	2018-05-29T13:10:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T12:45:00Z	2018-05-29T13:25:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T13:10:00Z	2018-05-29T13:50:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T13:25:00Z	2018-05-29T14:05:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T13:50:00Z	2018-05-29T14:30:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T14:05:00Z	2018-05-29T14:45:00Z	1	2018-06-04T17:00:00Z	IHC_1	10
SO_1_Batch	RSH_1	10	2018-05-29T08:30:00Z	2018-05-29T14:30:00Z	2018-05-29T15:10:00Z	1	2018-06-04T17:00:00Z	IHC_1	10

表 3.17 機台資訊紀錄表

Date	ToolID	ToolGroup	Run	Idel	Down	ChangeLine	UTIL
2018-05-30T07:59:59Z	IHC_1	IHC	0.361	0.035	0.604	0.010	0.9123
2018-05-31T07:59:59Z	IHC_1	IHC	0.391	0.005	0.604	0	0.9883
2018-06-01T07:59:59Z	IHC_1	IHC	0.303	0.093	0.604	0	0.7661
2018-06-02T07:59:59Z	IHC_1	IHC	0.000	0.000	0.000	0	0
2018-06-03T07:59:59Z	IHC_1	IHC	0.281	0.323	0.396	0	0.4651
2018-06-04T07:59:59Z	IHC_1	IHC	0.145	0.251	0.604	0	0.3655

將決策樹法則套用至表 3.17 機台資訊紀錄表來安排維修保養排程，可以得到新的機台開關機設定表，如表 3.18。

表 3.18 機台開關機設定表

Tool Name	Up	Down	AbsUp	AbsDown
IHC_1	8:00:00	17:30:00		
IHC_1			5/31/2018 17:30	5/31/2018 17:00
IHC_1			6/2/2018 17:30	6/2/2018 8:00
IHC_1			6/7/2018 17:30	6/7/2018 17:00
IHC_1			6/16/2018 17:30	6/16/2018 8:00

除了原本兩週一次固定維修保養額外新增了兩次根據決策樹當機法則安排的一般性維修保養，分別為 2018/5/31 17:00 至 2018/5/31 17:30 以及 2018/6/7 17:00 至 2018/6/7 17:30，將表 3.13 至表 3.15 機台群組表投入 CPS 模擬系統，觀察其產出的結果，表 3.19 為新增決策樹當機法則所安排的維修保養排程後，得到的機台資訊紀錄表。

表 3.19 新增決策樹當機法則之機台資訊紀錄表

Date	ToolID	ToolGroup	Run	Idel	Down	ChangeLine	UTIL
2018-05-30T07:59:59Z	IHC_1	IHC	0.465	0.035	0.500	0.010	0.9305
2018-05-31T07:59:59Z	IHC_1	IHC	0.791	0.005	0.204	0	0.9942
2018-06-01T07:59:59Z	IHC_1	IHC	0.585	0.093	0.322	0	0.8634
2018-06-02T07:59:59Z	IHC_1	IHC	0.000	0.000	0.000	0	0
2018-06-03T07:59:59Z	IHC_1	IHC	0.865	0.115	0.020	0	0.8825
2018-06-04T07:59:59Z	IHC_1	IHC	0.675	0.203	0.122	0	0.7688

### 3.3 績效比較階段

#### 3.3.1 決策評估

評估分類效能與一致性在歷史數據中上所找到的最佳法則，將資料分為訓練組及測試組，利用測試組資料觀察所得到的分類正確率，此正確率是否符合模型。並將其發展出來的法則整理成表格，如上面所提到表 3.8，若是依據法則歸納出來，則表格中 MTTR 維修時間設定為 30 分鐘，維修當機的機率利用決策樹的決策數量以及反決策數，計算其機率，考量維修當機發生的機率，此機率當進入排程時決定是否進行保養。

而使用者可以自行在決策樹維修策略表表中定義當機比率，即設定一定發生當機機率，超過這個比率時必須進行維修保養，發生機率(Probability)為使用者可自行設定之機率，若分類為短期當機時，使用者能自行決定是否將此列入時程規劃中。上述維修當機的機率未超過發生機率(Probability)所設定之機率，則不列入維修時程規劃中，能避免產能下降。

表 3.20 決策樹維修策略表-使用者設定

Machine Name	Amount	Upper_Bound	Amount	Lower_Bound	Time	Upper_Bound	Time	Lower_Bound	MTTR	Probability	DownType	CONT	N-cont

並透過模擬觀察表 3.17 及表 3.19 兩張機台資訊紀錄表，利用決策樹當機法則安排出維修保養排程與僅安排每兩週一次的維修保養排程比較兩種維修保養排程方法對機台資訊紀錄表之影響，可以明顯的發現兩張表之 Down time 比較下，表 3.19 的比例下降許多，機台突發當機比率降低以及機台稼動率提升。

## 第四章 實證研究

本章節將半導體晶圓廠套用此維護決策支援模組，針對其中當機資料進行討論，並進行實證研究。透過本研究所提出之三階段模組發展規則、模擬並進行績效比較。以下以各小節呈述之。

### 4.1 規則生成階段

本研究案例對象為 M 公司半導體晶圓廠之實際資料，以真實生產環境下所提供資料建立機台維修時程規劃之決策樹，並去驗證此模型是否符合真實狀況，利用 Weka 作為我們的求解工具，使用其中 J48(即 C4.5 決策樹)方法，從資料中找尋分類規則。

#### 1. 產品資訊：

針對關鍵機台群組所生產產品做分析，經過此機群之生產產品，計算所有經過此機群生產數量及生產加工時間。

#### 2. 保養資訊：

機台的相關資料來源由 MES 得知機台設備相關資料，並從 ERP 得知相關保養資訊。表 4.1 可以了解每個機台的加工時間、維修時間以及所屬機群。

表 4.1 機台維修

機台名稱	可加工時間 (min)	維修時間 (min)	機台群組
IHC06	120	30	IHC

#### 3. 機台資訊：

主要內容為其機台名稱及機台群組，並了解機台目前情況，因此有包含其狀態及當機、維修之相關資訊。依資料處理前後分為兩部分講解。

##### (1) 資料處理前：

本研究資料來自 MES、ERP 兩系統中，將所蒐集之現場機台生產資訊與機台使用狀態資料，各機台資料及生產數量彙整。

表 4.2 機台生產狀況

設備名稱	機台生產狀況	累積生產數量	狀態開始時間
IHC06	2	0	2013/01/01 15:18:45

(2)資料處理後:

本研究目的在於提供機台維修保養時程決策參考，投入決策樹前須先進行資料處理，處理完後以生產數量、加工時間、狀態為投入資料。投入資料包括以下:

N：正常運作

S：短期當機( $dt < 12$  小時)

M：中期當機( $12 < dt < 24$  小時)

L：長期當機( $dt > 24$  小時)

Amount：累積生產數量

Days：累積加工天數

表 4.3 投入資料

Status	Amount	Days
N(正常運作)	358	1
S(短期當機)	366	2
M(中期當機)	360	3
L(長期當機)	405	4

#### 4. 機台參數：

由於半導體製程複雜，設備成本昂貴，為了讓機台使用效率一直處於高效的運作狀態下，機台常會因為不同產品被分配在不同的機台群組當中，本研究將會針對關鍵機台群組作為對象。

#### 5. 分析系統之作業環境：

本研究將使用 Weka 軟體做分群(K-means)以及分析(C4.5 決策樹)，Weka 是一套完整的資料處理工具，包括對資料進行預處理、學習演算法、

分類、迴歸及資料視覺化的使用介面，使用者可以快速的方法處理資料，同時還能比較和評估不同的學習演算法的性能，能夠找到解決問題最有效的方法。

#### 6. 模擬軟體：

本研究使用之模擬軟體為 CPS 系統。特色為具有物件導向、離散事件、動態引擎等，使用 JAVA 程式語言做為主要系統開發工具，並結合 SQL 資料庫進行資料整合匯入，以 Excel 形式匯出結果。特點如下：

- (1) 動態模擬排程
- (2) 物件導向模型
- (3) 物件與報表管理
- (4) 使用 JAVA 程式語言作為邏輯開發
- (5) 使用 SQL 資料庫連結

### 4.1.1 資料預處理

本研究以半導體產業做為案例說明，原因是半導體生產設備成本相較其他大部分產業設備成本高，且如發生緊急狀況(如:當機)將會導致半導體業者的製造成本遽增，而在實務上，彈性維修比固定維修更符合實際狀況，若依造傳統的固定時間或經驗法則，並沒有依造實際現場生產設備之當機特性進行個性化維修保養，將無法有效提升整廠機台的生廠效率。

本研究將投入半導體十二吋晶圓廠的歷史資料，並以一般工廠會有的機台累積生產數量、機台累積生產天數、維修機台時間作為實驗因子，利用分群與決策樹，提出最適合的維修保養建議。透過資料探勘技術找尋過去維修保養資料，從中找到其分類，將半導體晶圓廠所提供之三個月實際數據如表 4.4，將原始資料依據 3.1.1 節介紹依據步驟進行資料預處理。

表 4.4 原始資料表

EqpID	E10	Amount	StateName	ReasonCode	Userld	StartTime
IHC06	2	0	SBY-Standby	[AL0001] OK	14542	2013/01/01 17:20:32.455622
IHC06	1	10	PRD-Work for	[AL0001]	81374	2013/01/01 19:56:19.606125
IHC06	3	0	TD experimen	[TS0008] down	00438	2013/01/01 20:00:19.606126
IHC06	2	0	SBY-Standby	C39206.0P from	TCS	2013/01/02 09:58:45.262326
IHC06	1	21	PRD-Work for	C39331.05 from	TCS	2013/01/02 10:30:01.404794
IHC06	2	0	SBY-Standby	C39331.05 from	TCS	2013/01/02 10:45:24.845501
IHC06	1	84	PRD-Producti	8A2438.00 from	TCS	2013/01/02 12:59:42.023425
IHC06	2	0	SBY-Standby	8A2438.00 from	TCS	2013/01/02 13:55:16.946394

透過實驗來評估本研究提出方法之績效，利用本研究之維修決策支援模組進行求解，希望透過不同參數設定進行比較，並分析本研究所提出之方法結果的績效及機台利用率。

本研究維修決策支援模組參數之實驗設計如下表 4.5 所示，對設立三種不同的實驗參數，分別為生產數量(2種)、訂單加工時間(2種)、機台數量(2種)，由以上參數進行實驗設計，用以測試不同情況下，準確度進行探討分析。

表 4.5 參數設計

實驗參數	value
生產數量	500、1000
訂單加工時間	7 天、14 天
機台數量	3 台、5 台

1. 情境一：

設定參數為機台數量與訂單加工天數，機台數量為 3 台與 5 台，數量定量為 500；訂單加工天數為 7 天、14 天。

將生產數量為 500 的情況下，來投入維修決策支援模組中做為比較，比較其正確率。可以由表 4.6 發現在機台數為 3 台、訂單加工天數為 7 天時，正確率為 72.19%，為四組數據當中最高的準確率；機台數為 5 台、訂單加工天數為 14 天時，正確率為 62.6%，為四組數據當中最底的準確率。

其中，可以發現當訂單加工天數為 7 天時比 14 天的準確率來的高，以及機台數為 3 台比 5 台的準確率來的高。

表 4.6 生產數量為 500 情況下

天數\機台數	3 台	5 台
7 天	72.19%	65.16%
14 天	69.08%	62.6%

## 2. 情境二：

設定參數為機台數量與訂單加工天數，機台數量為 3 台與 5 台，數量定量為 1000；訂單加工天數為 7 天、14 天。

將生產數量為 1000 的情況下，來投入維修決策支援模組中做為比較，比較其正確率。可以由表 4.7 發現在機台數為 3 台、訂單加工天數為 14 天時，正確率為 73.08%，為四組數據當中最高的準確率；機台數為 5 台、訂單加工天數為 7 天時，正確率為 63.18%，為四組數據當中最底的準確率。

其中，可以發現當訂單加工天數為 14 天時比 7 天的準確率來的高，以及機台數為 3 台比 5 台的準確率來的高。

表 4.7 生產數量為 1000 情況下

天數\機台數	3 台	5 台
7 天	64.57%	63.18%
14 天	73.08%	69.08%

由上述兩張表可以得知，在機台為 3 台以及生產數量為 1000，並且當訂單加工天數於 14 天時，決策支援模組可以得到最高的準確率為 73.08%。因情境模擬可得知機台、生產數量以及訂單加工天數的不同，會形成不同準確率的產生。因此，將準確率較高做為將此種類決策樹運用在實際案例中。

### 4.1.2 決策樹發展規則

透過上述資料處理與情境分析後，可得知機台為 3 台以及生產數量為 1000，並且當訂單加工天數於 14 天時，決策支援模組可以得到最高的準確率。因此由上述的因子，針對關鍵機群資料，投入維修決策支援模組中，以 80:20 比例，分為訓練資料及驗證資料，可得到如圖 4.1 所示的關鍵機台決策樹。



可以發現此決策樹正確率為 77.14%

```

=== Summary ===
Correctly Classified Instances      108           77.1429 %
Incorrectly Classified Instances     32           22.8571 %
    
```

圖 4.2 關鍵機台決策樹正確率

## 4.2 模擬結果階段

使用情境模擬後可知道相關因子的影響性，與其交叉比對的正確率，透過正確性確認決策樹的可行性後，並利用 CPS 模擬系統進行有無使用維修支援決策模組之法則，進行維修保養時程排程之績效驗證。將歸納出來之法則導入 CPS 模擬系統中之決策樹維修策略表，依據填入其機台、生產數量上限與下限、加工時間上限與下限以及維修時間等。透入因子後進行模擬，依照此法則進行維修保養之時程規劃。表 4.9 為機台 IHC01 之法則設定，MTTR 維修時間皆設為 30 分鐘。

表 4.9 案例決策樹維修策略表

Machine Name	Amount Upper_Bound	Amount Lower_Bound	Time Upper_Bound	Time Lower_Bound	MTTR
IHC01	250	200	6:00:00:00	4:00:00:00	0:00:30:00
IHC01	500	500	14:00:00:00	3:00:00:00	0:00:30:00
IHC01	900	800	3:00:00:00	1:00:00:00	0:00:30:00
IHC01	900	700	14:00:00:00	3:00:00:00	0:00:30:00
IHC01	700	650	6:00:00:00	4:00:00:00	0:00:30:00
IHC01	700	550	14:00:00:00	6:00:00:00	0:00:30:00
IHC01	1050	950	6:00:00:00	1:00:00:00	0:00:30:00
IHC01	1050	900	14:00:00:00	6:00:00:00	0:00:30:00
IHC01	1350	1150	6:00:00:00	0:00:00:00	0:00:30:00
IHC01	1100	1050	6:00:00:01	0:00:00:00	0:00:30:00
IHC01	1150	1100	6:00:00:02	0:00:00:00	0:00:30:00
IHC01		1350	14:00:00:00	7:00:00:00	0:00:30:00
IHC01	1400	1350	14:00:00:00	7:00:00:00	0:00:30:00
IHC01	1100	1050	10:00:00:00	1:00:00:00	0:00:30:00
IHC01	1350	1100	10:00:00:00	1:00:00:00	0:00:30:00
IHC01	1150	1100	14:00:00:00	7:00:00:00	0:00:30:00
IHC01	1100	1050	14:00:00:00	7:00:00:00	0:00:30:00

除了工廠所排定維修保養時間，當機台發生突發狀況時，可能會因機台過度使用，讓該機台較容易當機，或是機台可用時間降低，整條產線直接受到影響。表 4.10 為機台為期一個月內發生突發當機及工廠所排定維修保養。除了排定之維修保養外，可以發現突發當機次數高達九次，其中當機時間長達 1.5 天。

表 4.10 歷史資料機台當機

AbsUp	AbsDown	Status
2018/5/30 23:00	2018/5/30 11:00	5
2018/6/2 14:00	2018/6/2 11:00	5
2018/6/5 13:00	2018/6/4 11:00	5
2018/6/9 15:00	2018/6/9 05:00	5
2018/6/11 18:00	2018/6/11 10:00	4
2018/6/18 16:00	2018/6/18 09:00	5
2018/5/30 15:00	2018/5/30 11:00	5
2018/6/2 14:00	2018/6/2 11:00	5
2018/6/10 18:00	2018/6/9 11:00	5
2018/6/14 13:00	2018/6/14 02:00	4
2018/6/16 10:00	2018/6/16 09:00	5

使用維修支援模組，與原始當機機台的數據對照比較後，可以發現經由法則產生之數據分析與原始數據有六項資料符合，如表 4.11。其中，Status 狀態列顯示出可能會發生之當機狀況(例如:長、中、短當機)。

表 4.11 維修支援模組關鍵機台當機

AbsUp	AbsDown	Status
2018/5/30 11:30	2018/5/30 11:00	s
2018/6/2 11:30	2018/6/2 11:00	s
2018/6/9 05:30	2018/6/9 05:00	m
2018/6/18 09:30	2018/6/18 09:00	l
2018/6/2 11:30	2018/6/2 11:00	m
2018/6/9 11:30	2018/6/9 11:00	l

### 4.3 績效比較階段

經過原始數據與法則數據比對後，可得知經由法則所產生的預測高達六成，透過原始數據追查可以發現其中未預測到資料中，包含了停電等天然災害，無法避免。藉由表 4.12 和表 4.13 的兩張機台資訊紀錄表中當機比率進行比對，其中 Down 為一天中當機所占的比率，例如表 4.12 第一列 2018-05-30T07:59:59Z 原始 Down 為 0.5 代表 5-30 中有半天都處於當機狀態。

利用決策樹當機法則安排出維修保養排程與歷史數據的維修保養排程，比較兩種維修保養排程方法對機台資訊紀錄表之影響，可以明顯的發現機台突發當機比率降低。

表 4.12 IHC01 突發當機比率

DATE	原始Down	Down
2018-05-30T07:59:59Z	0.5	0.020833333
2018-05-31T07:59:59Z	0	0
2018-06-01T07:59:59Z	0	0
2018-06-02T07:59:59Z	0.125	0.020833333
2018-06-03T07:59:59Z	0	0
2018-06-04T07:59:59Z	0.874988426	0.874988426
2018-06-05T07:59:59Z	0.208344907	0.208344907
2018-06-06T07:59:59Z	0	0
2018-06-07T07:59:59Z	0	0
2018-06-08T07:59:59Z	0	0
2018-06-09T07:59:59Z	0.291678241	0.020833333
2018-06-10T07:59:59Z	0	0
2018-06-11T07:59:59Z	0.333333333	0.333333333
2018-06-12T07:59:59Z	0	0
2018-06-13T07:59:59Z	0	0
2018-06-14T07:59:59Z	0	0
2018-06-15T07:59:59Z	0	0
2018-06-16T07:59:59Z	0	0
2018-06-17T07:59:59Z	0	0
2018-06-18T07:59:59Z	0.291666667	0.020833333

表 4.13 IHC06 突發當機比率

Date	原始Down	Down
2018-05-30T07:59:59Z	0.166666667	0.166666667
2018-05-31T07:59:59Z	0	0
2018-06-01T07:59:59Z	0	0
2018-06-02T07:59:59Z	0.125	0.020833333
2018-06-03T07:59:59Z	0	0
2018-06-04T07:59:59Z	0	0
2018-06-05T07:59:59Z	0	0
2018-06-06T07:59:59Z	0	0
2018-06-07T07:59:59Z	0	0
2018-06-08T07:59:59Z	0	0
2018-06-09T07:59:59Z	0.874988426	0.020833333
2018-06-10T07:59:59Z	0.416678241	0
2018-06-11T07:59:59Z	0	0
2018-06-12T07:59:59Z	0	0
2018-06-13T07:59:59Z		0
2018-06-14T07:59:59Z	0.208344907	0.208344907
2018-06-15T07:59:59Z	0	0
2018-06-16T07:59:59Z	0.041666667	0.041666667
2018-06-17T07:59:59Z	0	0
2018-06-18T07:59:59Z	0	0

由表 4.12 與表 4.13 的數據彙整成圖 4.2 及圖 4.3，可明顯看出原始 Down 與經過維修決策支援模組 Down 的當機波動性。圖 4.2 紅色虛線為 IHC01 原始數據平均值、綠色為 IHC01 透過維修決策支援模組之數據平均值，可看出由維修決策支援模組後的結果，其當機平均值較低，有較好的結果預測；藍色線條為維修決策支援模組後的當機時間，可發現與橘色線條的當機結果有大部分相符，並且在波動性較大的時間(2018-06-13T07:59:59Z-2018-06-06T07:59:59) 時，有相同的預測結果。表示運用維修決策支援模組後，可得較好的運算結果，並能夠與原始數據的當機時間預測有相同結果。圖 4.3 紅色虛線為 IHC06 原始數據平均值、綠色為 IHC06 透過維修決策支援模組之數據平均值，可看出由維修決策支援模組後的結果，其當機平均值較低；橘色線條在波動性較大的時間(2018-06-08T07:59:59Z-2018-06-011T07:59:59) 時，藍色線條並無預測到，追蹤原始數據發現因突發狀況停電造成機台無法正常運作。

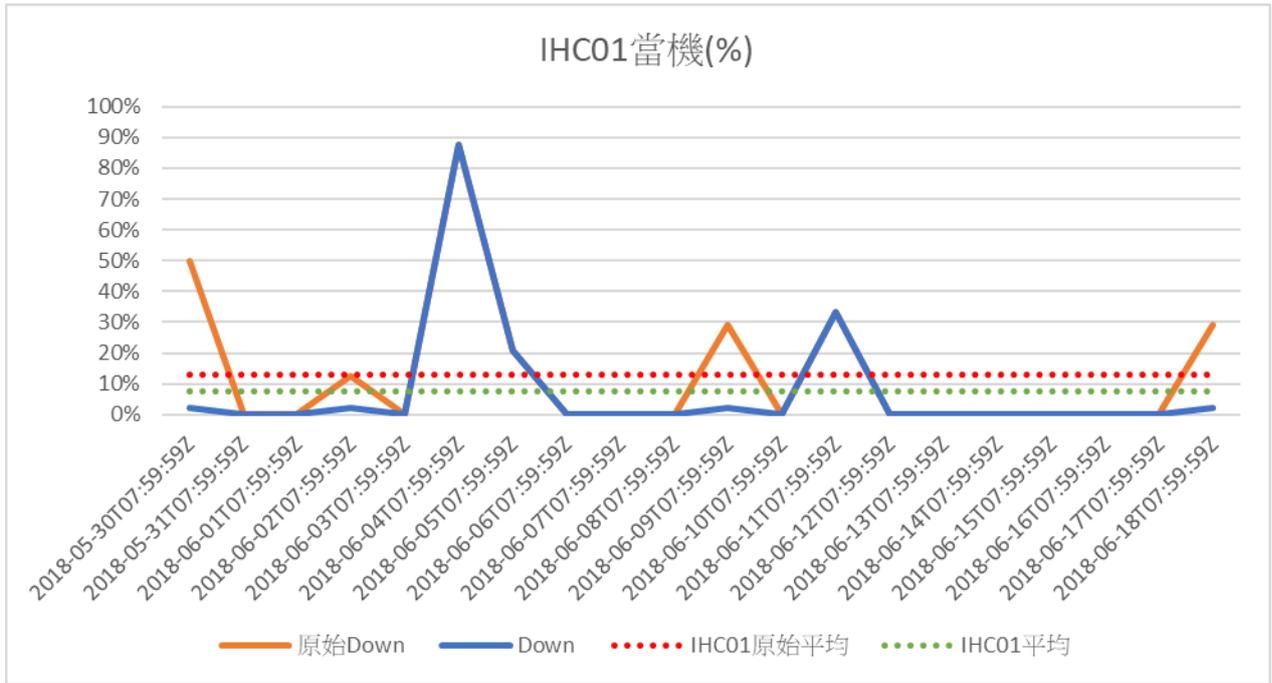


圖 4.2 IHC01 當機波動

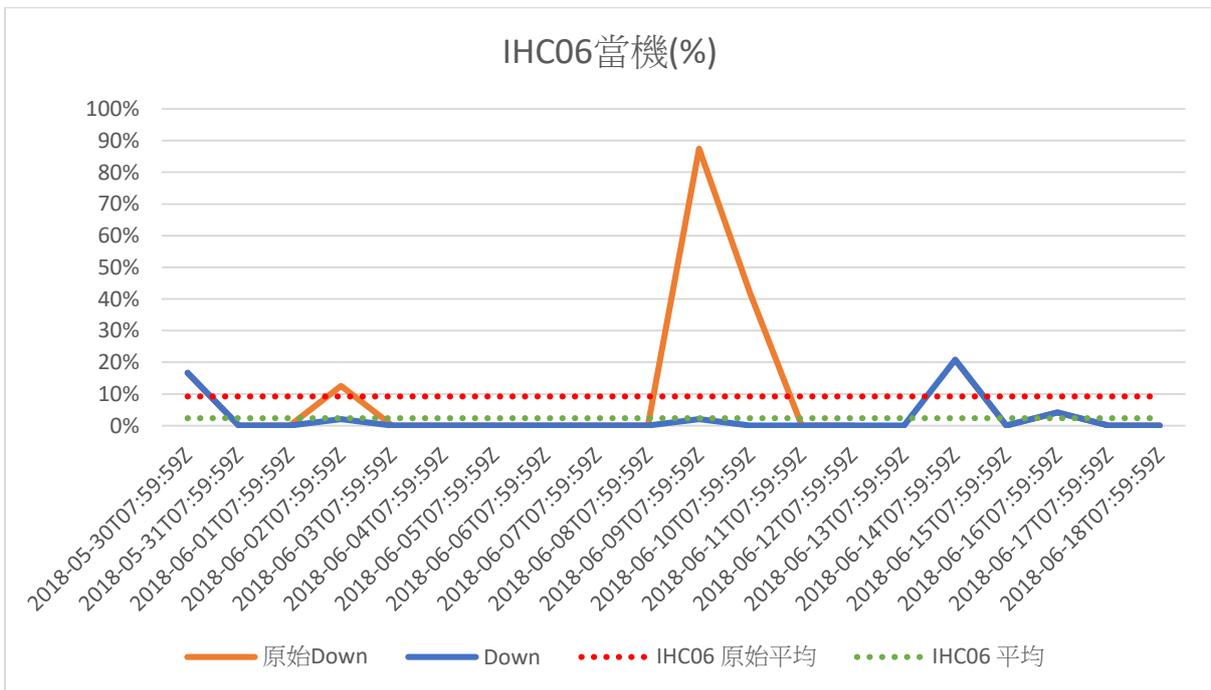


圖 4.3 IHC06 當機波動

將表 4.12 與表 4.13 的數據 Down 進行加總整理成表 4.14，可以發現透過維修決策支援模組的當機比率相對比歷史數據低，表示透過決策樹演算的結果，在可能發生長時間當機前，先安排維修保養，使當機時間有效縮短，可使機台使用率增加並提升產能。

本章利用維修決策支援模組排定維修時程規劃，並利用歷史數據、使用該模組的差異，以判斷維修決策支援模組能準確預測率達六成，且對於突發當機比率的改善有無實質效。

表 4.14 機台 Down 比率總和比較

	IHC01	IHC06
歷史數據	2.625	1.833
維修決策支援模組	1.5	0.458

## 第五章 結論與建議

### 5.1 結論

半導體產業建置時需大量的資金投入，設備昂貴，從實務經驗上，不定時發生的機台故障都將會對總體產能(或產出)有損失的影響，因此設備機台維護保養預測格外重要。如果不能妥善安排維護機台時間，機台容易發生當機情況，目前晶圓製造廠設備維修仍是設定為固定保養時間，每個廠不一樣，可能設定為兩週做一次設備的保養，而保養依賴資深維修人員的經歷，但設備更新快速、時間限制、人員流動等因素，造成技術傳承不易，當機台發生異常時，才會依照機台狀況進行故障診斷，但往往因為維修機台花費過多時間而造成損失。妥善規劃機台當機維修保養時程能確保機台正常運作、降低停機時間，成為影響各產業重要易議題之一。

本研究蒐集機台生產資料及狀態，大部分工廠維修保養針對定量保養及定時保養，針對這兩個因子及部分工廠所能收集到的資料，作為投入因子，且能符合大部分工廠，將維修支援模組導入至半導體 12 吋晶圓廠，以實際生產歷史資訊訓練出之決策樹發展出一套維修決策支援模組並搭配 CPS 模擬系統，每進行一次排程模擬及將維修決策支援模組之結果排入規劃中。將此法則投入模擬系統中進行排程並驗證其效益，利用維修決策支援模組安排出維修保養排程與僅安排每兩週一次的維修保養排程比較兩種維修保養排程方法對機台資訊紀錄表之影響，可以明顯的發現機台當機比率降低，可使機台使用率增加、稼動率提升並提升產能。

## 5.2 未來研究方向

本研究以半導體晶圓製造廠為基礎，考量機台生產累積數量及生產累積時間對於維修保養影響；然而在實際狀況所需考慮的變異並非僅此二因子，其可能受到電壓、機台內部物理性振動等影響，應有更多其他參數考量，其後續研究可將其他影響因子列入考量。

本研究建議之未來方向包含以下三點：

1. 執行決策支援模組，進行維修保養時程規劃時可納入更多變異因子。
2. 有更完整的資料進行參考，其投入因子更細分化生產數量、加工時間，原始資料的數據可看出更多現象。
3. 本研究目前針對關鍵機群，進行維修保養時程規劃，未來能進一步的應用本研究方法至其他機群中，以提供管理及排程人員更好的決策方法。

## 參考文獻

- [1]. 林思逸 (2015)。可用生產時間波動性對產能影響性分析-以半導體晶圓製造為例 (碩士論文)。東海大學，臺中市。
- [2]. 袁梅宇 (2016)。王者歸來：WEKA 機器學習與大數據聖經。台北市:佳魁資訊。
- [3]. 黃允成、謝雅惠、邱士珍 (2017)。多類型故障型態下最適預防保養策略研究。技術學刊，32(1)，13-38。
- [4]. 黃嘉莉、葉怡芬、許瑛瑄、曾元顯 (2017)。取得中學教職的關鍵因素:運用決策樹探勘師資培育歷程。教育科學研究期刊，62(2)，89-123。
- [5]. 簡禎富、李培瑞、彭誠湧 (2003)。「半導體製程資料特徵萃取與資料挖礦之研究」，資訊管理學報，10(1)，63-84。
- [6]. 簡禎富、蕭禮明、王興仁(2004)。建構半導體製造管理目標層級架構與製造資料之資料挖礦。工業工程學刊，21(4)，313-327。
- [7]. 簡禎富、林昀萱、鄭仁傑(2008)。「建構模糊決策樹及其在有交互作用之半導體資料之資料挖礦以提昇良率之研究」。品質學報，15(3)，193-210。
- [8]. Barkia, H., Boucher, X., Le Riche, R., Beaune, P., Girard, M. A., & Rozier, D. (2013). Semiconductor yield loss' causes identification: A data mining approach. In *Industrial Engineering and Engineering Management (IEEM), 2013 IEEE International Conference on* (pp. 843-847).
- [9]. Bose, I., & Mahapatra, R. K. (2001). Business data mining—a machinelearning perspective. *Information & Management*, 39, 211–225.
- [10]. Capodiec, L. (2017). Data analytics and machine learning for design-process-yield optimization in electronic design automation and IC semiconductor manufacturing. In *Semiconductor Technology International Conference (CSTIC), 2017 China* (pp. 1-3). IEEE.
- [11]. Casali, A., & Ernst, C. (2012). Discovering correlated parameters in semiconductor manufacturing processes: A data mining approach. *IEEE Transactions on Semiconductor Manufacturing*, 25(1), 118-127.
- [12]. Chauhan, D., & Jaiswal, V. (2016). An efficient data mining classification approach for detecting lung cancer disease. In *Communication and Electronics Systems (ICCES), International Conference on* (pp. 1-8). IEEE.
- [13]. Chien, C. F., Wang, W. C., & Cheng, J. C. (2007). Data mining for yield enhancement in semiconductor manufacturing and an empirical study. *Expert Systems with Applications*, 33(1), 192-198.
- [14]. Doostparast, M., Kolahan, F., & Doostparast, M. (2014). A reliability-based approach to optimize preventive maintenance scheduling for coherent systems. *Reliability*

- Engineering & System Safety*, 126, 98-106.
- [15].Ebrahimipour, V., Najjarbashi, A., & Sheikhalishahi, M. (2015). Multi-objective modeling for preventive maintenance scheduling in a multiple production line. *Journal of Intelligent Manufacturing*, 26(1), 111-122.
- [16].Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- [17].Jayaram, A., & Singal, S. (2017). An enterprise resource management model for business intelligence, data mining and predictive analytics. In *Cloud Computing, Data Science & Engineering-Confluence, 2017 7th International Conference on* (pp. 485-490). IEEE.
- [18].Jantan, H., Hamdan, A. R., & Othman, Z. A. (2011). Data mining classification techniques for human talent forecasting. In *Knowledge-Oriented Applications in Data Mining*. InTech.
- [19].Lie, C. H., & Chun, Y. H. (1986). An algorithm for preventive maintenance policy. *IEEE Transactions on Reliability*, 35(1), 71-75.
- [20].Liu, J., Yang, F., Wan, H., & Fowler, J. W. (2011). Capacity planning through queueing analysis and simulation-based statistical methods: a case study for semiconductor wafer fabs. *International Journal of Production Research*, 49(15), 4573-4591.
- [21].Munirathinam, S., & Ramadoss, B. (2014). Big data predictive analytics for proactive semiconductor equipment maintenance. In *Big Data (Big Data), 2014 IEEE International Conference on* (pp. 893-902). IEEE.
- [22].Munirathinam, S., & Ramadoss, B. (2016). Predictive models for equipment fault detection in the semiconductor manufacturing process. *IACSIT International Journal of Engineering and Technology*, 8(4), 273-285.
- [23].Nikolaidis, A., Asselin, S., Auster, M., & Bragg, N. (2013). Capacity planning for dynamic inter-data center networking via Erlang modeling. In *Optical Communication (ECOC 2013), 39th European Conference and Exhibition on* (pp. 1-3). IET.
- [24].Patidar, P., & Tiwari, A. (2013). Handling missing value in decision tree algorithm. *International Journal of Computer Applications*, 70(13), 31-36.
- [25].Romauch, M., & Hartl, R. F. (2017). Capacity planning for cluster tools in the semiconductor industry. *International Journal of Production Economics*, 194, 167-180.
- [26].Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27.
- [27].Sharma, A., Yadava, G. S., & Deshmukh, S. G. (2011). A literature review and future perspectives on maintenance optimization. *Journal of Quality in Maintenance Engineering*, 17(1), 5-25.
- [28].Tang, T. J., Chung, A., Zhao, A., Kang, R., Zhang, M., Chien, K., ... & Zhang, J. (2016, October). An IoT inspired semiconductor Reliability test system integrated with data-mining applications. In *Cloud Computing and Internet of Things (CCIOT), 2016 2nd International Conference on* (pp. 111-114). IEEE.

- [29]. Thuraisingham, B. (2014). *Data mining: technologies, techniques, tools, and trends*. Boca Raton: CRC press.
- [30]. Tso, G. K., & Yau, K. K. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9), 1761-1768.
- [31]. Wang, L. C., Wang, A., & Chueh, C. Y. (2016). Development of a capacity analysis and planning simulation model for semiconductor fabrication. *The International Journal of Advanced Manufacturing Technology*, 1-16
- [32]. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Burlington: Morgan Kaufmann.
- [33]. Yan, X., & Zheng, L. (2017). Fundamental analysis and the cross-section of stock returns: A data-mining approach. *The Review of Financial Studies*, 30(4), 1382-1423.
- [34]. Yu, C. M., & Kuo, C. J. (2016). Data mining approaches to optimize the allocation of production resources in semiconductor wafer fabrication. In *Semiconductor Manufacturing (ISSM), 2016 International Symposium on* (pp. 1-4). IEEE