中文摘要

資訊爆炸的時代，投資者容易因過量的資訊而產生混淆。過去研究表明，企業營運績效不佳時容易於發布之財務報表中採用較為艱深的詞彙以誤導投資者的判斷。而投資者若要擷取財務報表中之摘要則需要耗費大量時間。財務報表經常作為投資者評估企業的依據，且財務數值是以往較常用來評估企業營運表現的方法，然而，此種方法可能忽略財務報表中敘述型文字所隱含的重要資訊。半導體產業作為台灣的產業支柱，對台灣之重要性不言而喻，其績效表現自然會受到投資者格外關注。

因此，本研究採用 2020 年至 2022 年所有半導體產業上市公司做為研究樣本，旨在挖掘以及探討半導體產業發布財務報表之敘述文字所隱含的潛在風險，透過自然語言處理技術 Sentence BERT 語意分析模型進行文本向量的轉換、K-means 集群分群、風險文字分析以識別和觀察敘述型資料所潛藏的風險，並且利用企業營運績效分析工具之情境相依資料包絡分析法，將企業財務表現進行分析，而後將風險變數及財務變數經由 XGboost 模型預測其對企業績效之影響，研究結果顯示企業對風險的控制和財務表現皆對企業績效有強健的影響力。研究結果表明 2020 年至 2022年之半導體產業以宏觀經濟風險、資安風險以及訴訟風險為主要企業面臨之挑戰。透過此研究可以幫助企業維持或提升績效表現以及規避潛在之企業風險，同時也能讓企業其利害關係人更加了解企業的營運表現。

關鍵詞：風險管理、資料包絡分析法、Sentence BERT 語意分析模型、K-means 集群分群、XGboost 模型預測

ABSTRACT

In the era of information explosion, investors can easily become confused by the overwhelming amount of information available. Previous studies have shown that companies with poor operational performance tend to use more complex terminology in their financial reports to mislead investors' judgments. Extracting summaries from financial reports requires significant time and effort. Financial reports are often used by investors to evaluate companies, and financial metrics have traditionally been a common method for assessing business performance. However, this approach may overlook important information embedded in the narrative sections of financial reports. The semiconductor industry, as a cornerstone of Taiwan's economy, naturally draws significant attention from investors regarding its performance.

Therefore, this study uses all listed semiconductor companies from 2020 to 2022 as the research sample, aiming to uncover and explore the potential risks hidden in the narrative texts of financial reports published by the semiconductor industry. By employing natural language processing techniques with the Sentence BERT semantic analysis model for text vector conversion, K-means clustering, and risk word analysis, we identify and observe the latent risks in narrative data. Additionally, using the context-dependent Data Envelopment Analysis (DEA) tool, we analyze companies' financial performance. Subsequently, we predict the impact of risk variables and financial variables on business performance through the XGBoost model. The results indicate that both risk control and financial performance have a strong influence on corporate performance. The findings reveal that from 2020 to 2022, the semiconductor industry primarily faced macroeconomic risks, cybersecurity risks, and litigation risks. This study can assist companies in maintaining or enhancing performance and mitigating potential risks, while also helping stakeholders better understand the companies' operational performance.