

摘要

公司營運中，發生產品缺貨時會對公司造成相當大的影響，如：違約、賠償等，導致成本與人力的增加，因此公司庫存管理相當重要，隨著大數據與相關技術的興起機器學習被廣泛應用，且過去研究顯示機器學習有效進行庫存控制。一般而言，公司產品缺貨與非缺貨數量差距相當大，造成資料不平衡，使機器學習分類預測結果相當不準確，因此，本研究透過 Kaggle 競賽 Can You Predict Product Backorders? 材料缺貨數據集作為研究資料，分別使用 Tomek Links 欠採樣、SMOTE 過採樣、SMOTE-Tomek 混合採樣、SMOTE-ENN 混合採樣四種採樣方法，搭配隨機森林與 XGBoost 兩種機器學習演算法進行模型建立，使用 AUC、Recall、Precision、F-Measure 做為模型績效評估指標。結果顯示混合採樣方法搭配機器學習演算法有效提升產品缺貨分類的準確率。

關鍵字:存貨管理、資料不平衡、機器學習、隨機森林、XGBoost