**食品大數據智能監控研究**

**一、專案概述**

透過歷史查驗資料，以機器學習方法針對各產品中分類建立「邊境報單風險預測模型」，預測進口報單之不合格風險，做為業管單位決定是否抽驗之依據。

建模因子：六大面向如

因子處理、因子篩選、資料不平衡處理

演算法：

集成方式、驗證集挑選模型

**二、工作項目**

為專案團隊組員，負責組長(專員)交辦資料分析事項。

評估蔬果類之產品中分類模型訓練組資料保留期間，以農藥殘留法規更動時間點做設定。結果顯示調整資料保留期間對於提升模型預測表現無太大幫助，後續將評估結果供團隊參考，並納入做為模型年度大更新考量之參數。

註1：蔬果類之產品中分類含生鮮冷藏冷凍蔬菜(FC11)、生鮮冷藏冷凍水果(FC21)、香辛料(FC34)

註2：以F1-score評估模型預測表現(閾值0.5)。

測試不同產品中分類模型加入黑名單因子，是否能提升預測表現。結果顯示多數模型在加入黑名單產品種類或進口商時，預測表現有所提升，後續則將前述因子納入建模。

註：黑名單因子包含貨品號列、產品種類、生產國、進口商，將不合格率排序前25%之標的列為黑名單。

資料不平衡處理、數據分箱

**食品鏈**

**一、專案概述**

透過業者申報交易資訊，串接食品交易流向，掌握業者上下游關係，以強化食品追溯追蹤效能。當食安事件發生時，能迅速掌握問題食品流向，降低對民眾造成之危害。

註：依食品類別分別建構交易網路，食品類別含食用油脂、肉品加工食品(豬、牛、羊、禽)、乳品加工食品等。

**二、工作項目**

為專案(唯一)資料分析師，負責檢視系統問題與資料分析事項。

自行串接資料與廠商系統資料進行比對後，總結幾項較大的問題：

1. 嬰兒與較大嬰兒配方食品交易資料有缺漏情形，原因為漏接販售業之申報資料。
2. 未將標記為「無收貨」之資料排除，導致交易資料虛增。

註：為確保業者當月確實沒有收貨行為而非漏報，業管單位仍會要求業者申報「無收貨」。

1. 黃豆製品、糖(大宗物資)資料篩選邏輯有疑慮，原材料不應僅用簡單關鍵字(黃豆、糖)篩選。
2. 未參考官方公告之「食品輸入類別(非追產品分類)與貨品號列對照表」定義資料範圍，導致資料較為雜亂或有缺漏情形。
3. 由於交貨資料合併申報問題(交給數家業者合併總量申報)，須推算各下游業者之購買量。然計算上未考量其他交貨方式，如賣給消費者或自行使用，導致推估結果可能有過高之情形。

利用食品業者上下游交易資料建立交易網路(依產品類別區分)，找出網路中相對重要業者(依特徵向量中心性)，提供業管單位做為政策推動輔導規劃對象

因應美豬進口議題，串接進口資料與交易流向，提供業管單位做儀表板監控使用

**逾期風險監控**

透過統計方法，尋找可能從事逾期食品違法行為潛在風險業者

後市場

化學物質流向