

機器學習-期中報告

銷售數量預測

學號：5111029028

學生：劉廷恩

目錄

一、設定研究目標

二、設定假設

三、資料收集

四、驗證假設

五、檢定假設

六、參考資料

一、設定研究目標

以中部某製造公司為例，由該公司是屬於金屬加工製造加工類型的製造業，其生產產品類別分類較為繁雜，主要可分為醫療手術用品、精密扣件、微波開關...等分類，該公司銷售模式主要是為 B2B(公司對公司交易)。

故如可先預測未來一年的各類型產品的銷售數量則可先進行生產排程作業，達到降低臨時作業產生成本與違約損失(員工加班費、出貨數量未達成...等)。

二、設定假設

設定的特性值(feature)可分為內部、外部資料，進行銷售數量的預測，內部資料為公司內部紀錄資料，外部資料為市場環境資料，故透過日期、物料號碼、客戶編號、去年本月平均匯率、前期月平均匯率...等資料進行銷售數量預測。

三、資料收集

參考論文(以機器學習建構一外銷銷售預測模型之研究，2018，蔡雅婷，國立屏東大學)研究方式將資料分為內部與外部資料。

內部資料主要為公司銷售資料，資料來源透過企業資源規劃系統(簡稱為ERP)內銷售業務所產生的歷年銷售資料紀錄，資料初步抓取筆數約 16萬筆，需再進行資料篩選與處理(排除服務性料號、排除退貨資料、排除無效資料、日期處理...等)。

1. 物料號碼

2. 出貨數量
3. 客戶編號
4. 出貨日期，需再轉換為年、月
5. 交易幣別
6. 客戶風險資料
7. 產品階層
8. 物料群組
9. 單位
10. 交貨種類
11. 銷售組織

外部資料主要為外部市場環境資料，資料來源透過經濟部統計資料。

1. 客戶使用貨幣與台幣該期間交易月平均匯率
2. 歷年通貨膨脹率

四、驗證假設

參考論文(以機器學習建構一外銷銷售預測模型之研究，2018，蔡雅婷，國立屏東大學)研究結果是以 Random Forest、Logistic regression、Neural Network 模型進行分析。

由於需要進行數量預測，所以需計算的 Y 需為數值類型，依造目前學習的方式，我認為 Linear regression 相較於 Logistic regression(分類)更為適合，故將 Logistic regression 替換為 Linear regression。

故使用 Random Forest、Linear regression、Neural Network。

五、檢定假設

由於資料預計抓取區間為 2019/09 至 2022/12。

2023 年一月至十二月每月依造模型產出預測結果除以 2023 年一月至十二月每月實際出貨數量之絕對值，其值需小於 5 %。

六、參考資料

1. [以機器學習建構一外銷銷售預測模型之研究，2018，蔡雅婷，國立屏東大學](#)
2. [經濟部統計匯率資料](#)
3. [歷年物價指數](#)
4. [通貨膨脹率](#)