

廣告效益提升方案： A/B測試與機器學習實踐

Machine Learning
Random forest
A/B testing
Seaborn
Python
Pandas
EDA

Machine Learning
Random forest
A/B testing
Seaborn
Python
Pandas
EDA

相關資料：<https://reurl.cc/80zabX>

林哲緯
Jared Lin

目錄

- 專案概述.....3
- 探索式資料分析（ EDA ）6
- A/B測試分析.....9
- 模型建立與應用.....11
- 結論與建議.....13

專案概述

通過A/B測試和機器學習模型的建立與應用，分析和預測不同廣告策略的成效，從而為制定更加精準有效的市場營銷策略提供數據支持。

- 廣告投放的效益？（觀看廣告轉化至購買）
- 廣告投放的時機？（預測最佳的投放方式）

專案概述

【資料來源】

- Kaggle 公開資料集
(<https://www.kaggle.com/datasets/faviovaz/marketing-ab-testing>)

【資料共588101筆，重要欄位有】

- 觀看廣告 / 觀看非廣告
- 購買產品 / 未購買產品
- 廣告數量、時間

專案概述

【相關套件和技術】

- Pandas, NumPy：數據處理、運算和分析。
- Matplotlib, Seaborn：數據視覺化和圖表製作。
- 卡方檢驗 (Chi-squared test)：統計檢驗，用於A/B測試的結果分析。
- Scikit-learn：機器學習模型的建立和評估。
- SMOTE (imbalanced-learn)：處理類別不平衡問題，提高少數類樣本的預測性能。
- RandomForestClassifier：分類問題的隨機森林模型。
- GridSearchCV：參數優化和模型選擇。

探索式資料分析

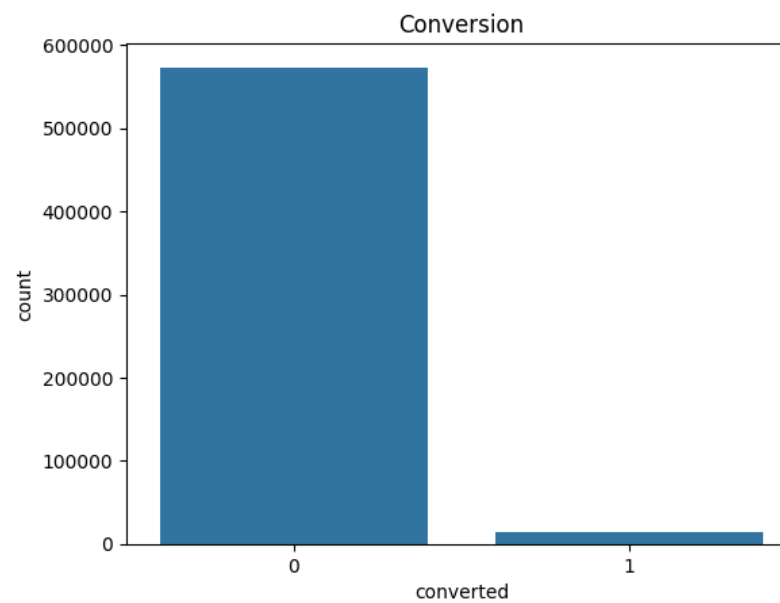
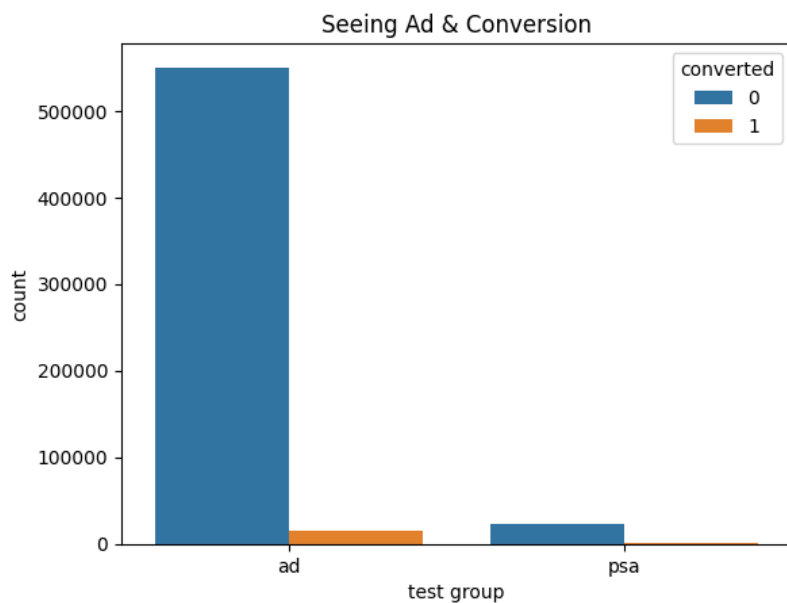
【資料簡述】

- test group：觀看廣告為ad；觀看其他非廣告內容為psa
- converted：購買產品為True；未購買產品為False
- total ads：觀看廣告的次數
- most ads day：在一週中哪天觀看最多廣告（星期）
- most ads hour：在一天中哪個時段觀看最多廣告（24小時）

	Unnamed: 0	user id	test group	converted	total ads	most ads day	most ads hour
0	0	1069124	ad	False	130	Monday	20
1	1	1119715	ad	False	93	Tuesday	22
2	2	1144181	ad	False	21	Tuesday	18
3	3	1435133	ad	False	355	Tuesday	10
4	4	1015700	ad	False	276	Friday	14

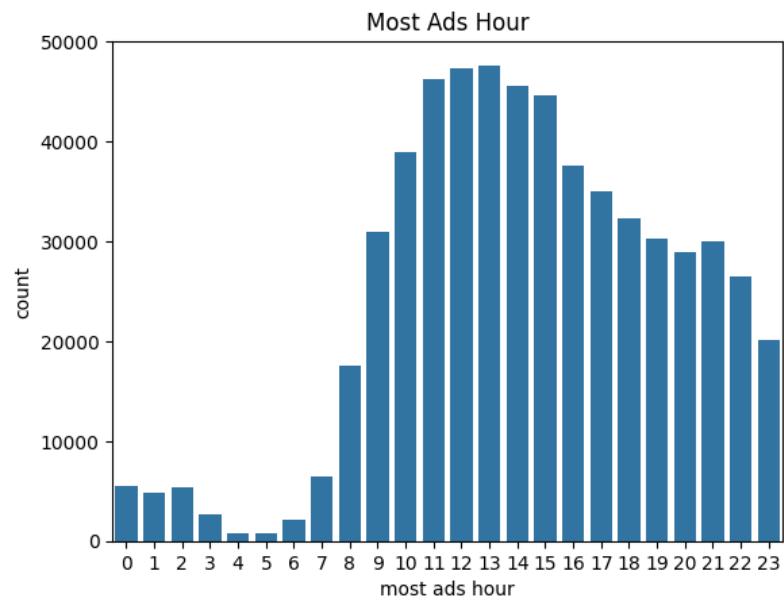
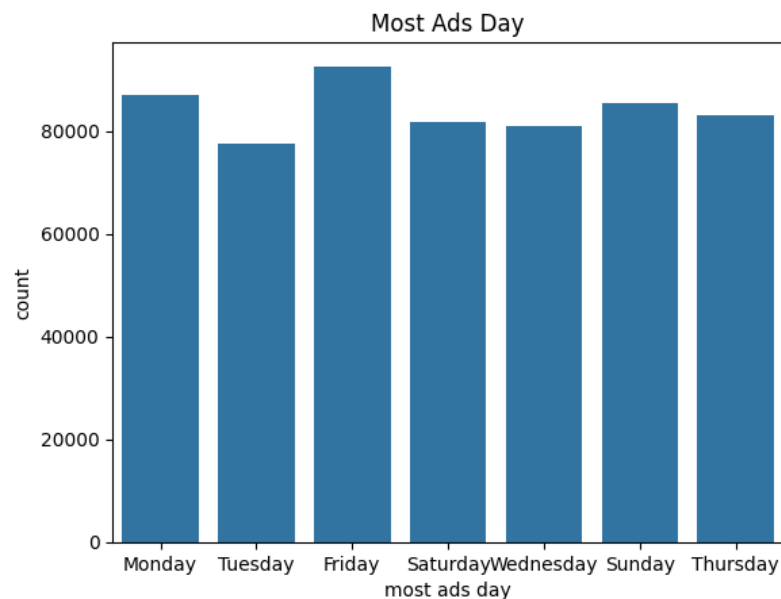
探索式資料分析

- 觀看廣告與否的差異大
 - 購買產品與否的差異大
- ⇒ 觀看廣告者是否有助於增加購買產品的可能 (a/b testing)
- ⇒ 能否有效以廣告投放預測購買產品 (machine learning)



探索式資料分析

- 一週中觀看廣告數平均
 - 一天中廣告數集中在中午至下午
- ⇒ 探討廣告投放時間與購買產品之關係（模型預測結果）



A/B 測試分析

【假說】

- 零假設 (H_0) : 廣告投放對轉換率無影響。
- 對立假設 (H_1) : 廣告投放能顯著提高轉換率。

【結論】

- 廣告對於提高用戶的轉換率具有顯著的正面影響。

A/B 測試分析

【數據】

- χ^2 (卡方值) : 54.01 ; p_value (P值) : $1.999e-13$
- expected (期望頻率) :

看到廣告：期望非轉換數550327.72，期望轉換數14249.28

未看廣告：期望非轉換數22930.28，期望轉換數593.72

【初步建議】

- 廣告策略調整：根據測試結果，建議重點關注並優化廣告內容和投放策略，進一步提升轉換效率。

模型建立與應用

【模型】

- 透過觀看廣告與否、廣告數、廣告時間等資料，預測使用者的轉換率（是否購買產品）
- 使用隨機森林（Random Forest）算法來尋找最佳模型，並通過網格搜索（GridSearchCV）來優化模型參數

模型建立與應用

【結果】

- 根據新的使用者資料，可以預測使用者是否購買產品
- 準確率：81.29%
- 召回率：76%

結論與建議

【策略應用】

- 精準客戶定位：通過分析預測數據，能夠鑑定具有最高轉化潛力的用戶群體。建議利用這些見解，執行更為個性化的營銷策略，對這些群體進行定向廣告投放，從而最大化營銷預算的效益和提高總體ROI。

結論與建議

【模型演進】

- 數據增強與模型迭代：為了持續提升模型的準確性和適用性，應該定期將新收集的數據（如用戶反饋、廣告互動情況）納入訓練集中。
- 可以嘗試將廣告內容的具體元素（如語言風格的差異）作為特徵引入模型，這將有助於進一步理解不同廣告元素對用戶行為的影響。

廣告效益提升方案： A/B測試與機器學習實踐

Machine Learning
Random forest
A/B testing
Seaborn
Python
Pandas
EDA

Thank You

相關資料：<https://reurl.cc/80zabX>

林哲緯
Jared Lin