



利用機器學習預測 iPhone與Samsung手機未來的市占率

研究動機與目的

隨著智慧型手機技術的快速發展，全球市場競爭加劇，iPhone與Samsung作為兩大領導品牌，始終在市場中佔據重要位置。技術創新、顧客滿意度及品牌忠誠度在這場競爭中扮演著關鍵角色。而隨著消費者需求的變化及新興市場的崛起，了解這些因素如何影響iPhone與Samsung的競爭模式變得尤為重要。因此，我們將深入探討兩種品牌在市場上的競爭動態，揭示未來的發展趨勢。

研究範圍與架構

- 在國際數據公司(IDC)中2015年至2024年間iPhone與Samsung智慧型手機之數據
- 變數數據為Y=市占率(%),
X=銷售量、價格、品牌忠誠度、顧客滿意度



研究方法

(1) 移動平均法

我們將利用移動平均法所預估出的銷售量、價格、品牌忠誠度及顧客滿意度的未來值，來當作未來市占率的變數預測，並接著使用線性迴歸及隨機森林方法。

(2) 線性迴歸

在Python程式碼上我們將使用scikit-learn函式庫中的LinearRegression 函式來建立線性迴歸模型，並對市占率問題進行預測。

(3) 隨機森林

在Python程式碼上，scikit-learn中也提供了隨機森林迴歸的模型，我們會使用函式庫中RandomForestRegressor函式來執行隨機森林迴歸，並對問題進行預測。

研究結果

1. 資料異常值測試

```
異常值:
Empty DataFrame
Columns: [品牌, 年份, 季度, 銷售量 (萬台), 價格, 品牌忠誠度, 顧客滿意度, 市佔率 (%) ]
Index: []
```

給定的數據中沒有發現異常值

2. 建立線性迴歸模型的特徵係數

iPhone 的線性迴歸特徵係數:			Samsung 的線性迴歸特徵係數:		
	特徵	係數		特徵	係數
0	銷售量 (萬台)	2.139632	0	銷售量 (萬台)	1.129920
1	價格	0.744568	1	價格	-0.819150
2	品牌忠誠度	0.890161	2	品牌忠誠度	-0.398614
3	顧客滿意度	-0.726293	3	顧客滿意度	0.605906

兩者的銷售量 (萬台) 是正向影響力最大的特徵，增加它們會顯著的提升市占率。

4. 隨機森林

① 使用 sklearn 函式庫中的LinearRegression 以建立線性迴歸模型

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
```

② 建立隨機森林模型，並與3.線性迴歸的③和④步驟相同，以預測隨機森林的未來值

```
rf_model = RandomForestRegressor(n_estimators=100,
                                max_depth=None, random_state=42)
rf_model.fit(X_train_scaled, y_train)
```

3. 線性迴歸

① 使用 sklearn 函式庫中的 LinearRegression 以建立線性迴歸模型

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

② 建立線性迴歸模型

```
model = LinearRegression()
model.fit(X_train_scaled, y_train)
```

根據線性迴歸的預測公式，其中Y是預測的市佔率(%)

iPhone為： $Y=1.4827+2.1396 \cdot x_1+0.7446 \cdot x_2+0.8902 \cdot x_3-0.7263 \cdot x_4$

Samsung為： $Y=1.4827+1.1299 \cdot x_1-0.8192 \cdot x_2-0.3986 \cdot x_3+0.6059 \cdot x_4$

x_1 是銷售量， x_2 是價格， x_3 是品牌忠誠度， x_4 是顧客滿意度

③ 建立線性迴歸模型的性能評估，均方誤差和R²分數

```
# 模型性能評估
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f"模型評估 - MSE: {mse:.4f}, R²: {r2:.4f}")
```

④ 代入移動平均法所算出來的 iPhone與Samsung銷售量、價格、品牌忠誠度及顧客滿意度所估出來的未來值

```
df = pd.read_csv('new_data.csv')
# === 定義未來數據 ===
future_data_iphone = pd.DataFrame({
    '銷售量 (萬台)': [5876],
    '價格': [29900],
    '品牌忠誠度': [90],
    '顧客滿意度': [81]
})
future_data_samsung = pd.DataFrame({
    '銷售量 (萬台)': [5363],
    '價格': [28480],
    '品牌忠誠度': [75],
    '顧客滿意度': [80]
})
```

⑤ 預測結果

iPhone 的預測結果:
模型評估 - MSE: 4.5583, R²: 0.5374
iPhone 未來市佔率預測: 17.71 %

Samsung 的預測結果:
模型評估 - MSE: 0.8234, R²: 0.4060
Samsung 未來市佔率預測: 18.76 %

⑤ 預測結果

iPhone 的預測結果:
模型評估 - MSE: 9.2675, R²: 0.0595
iPhone 未來市佔率預測 (Random Forest): 19.83 %

Samsung 的預測結果:
模型評估 - MSE: 0.6423, R²: 0.5366
Samsung 未來市佔率預測 (Random Forest): 19.29 %

5. 模型比較

由橘色圈圈得知根據R²定理，線性迴歸模型對iPhone的整體表現更好，相較於隨機森林更接近於1，對未來市占率的預測也較合理。

由紅色圈圈得知根據MSE定理，隨機森林模型對Samsung的預測誤差更小，對未來市占率的預測較合理。

結論

再經過模型比較後，線性迴歸模型在解釋 iPhone 市佔率變化的能力較強，而隨機森林模型在處理 Samsung 數據時效果更佳

iPhone的未來市占率為：17.71%；Samsung的未來市占率為：19.29%

未來方向

- 提升模型準確性與穩定性
- 擴展研究範疇與應用場景
- 融入更多動態數據
- 強化數據多樣性與規模
- 跨領域合作與應用
- 自動化與即時分析平台的開發