**國立金門大學資訊工程學系專業實習系統功能檢測彙整表**

|  |  |  |
| --- | --- | --- |
| **組 別：第六組** | | **指導老師：張珀銀** |
| **組 名：建構基於機器學習與時間序列方法的加密或幣分析平台** | | **學生：**  **110910537王鴻銘**  **110910521張展睿** |
| **第一次填單日期：111/11/12**  **第二次填單日期：112/01/01**  **第二次填單日期：112/04/20** | |
| 系統  功能  規劃  說明 | **1.預計在學期內完成的功能與規格說明: (最少3項)**  1. 平台前端設計  2. 平台後端設計  3. 機器學習模型訓練評估  4. 機器學習模型效能分析  5. 整合平台模型API  **2. 所需投入人力時間與設備等如附件說明 (必填)**  3**. 詳細系統架構與演算流程等如附件說明 (必填)** | |
| 單元  功能  測試  實驗  說明 | 1. **功能與規格測試結果說明: (最少3-5項)**   1. 平台前端設計 100%  2. 平台後端設計100%  3. 機器學習模型訓練評估100%  4. 機器學習模型效能分析100%  5. 整合平台模型API 100%  **2. 詳細實驗功能測試結果與效能如附件 (必填)** | |
| 單元  功能  改進  與  單元  介面  整合  說明 | **功能與規格測試檢討改進與下次預定進度規劃 (必填)**  1. 時空分析模型訓練評估  2. 時空分析模型效能分析  3. 平台前端開發  4. 平台後端開發  5. 撰寫成果報告書 | |
| 老師  評語 | **評估可行性及其後續處理事項：**  **□優良(80-99) □待改善(65-79) □極待加強 (51-64) □不通過 (40-50)** | |
| 主任  評語 | **評語:**  **□優良(80-99) □待改善(65-79) □極待加強 (51-64) □不通過 (40-50)** | |
| 備註 | **(1)每組每學期開學第三週前交學期規劃或是延續上學期功能延伸的專案規劃說明。**  **(2)請每組依進度每學期交3次檢測表，依學期規劃進度撰寫。**  **(3)請先寫預計在一年內完成專案的規劃期程，並於規劃內完成專案報告。**  **(4)需改整個專題主題請於系統功能規劃說明原因與目的。**  **(5)各組期末考前2週將資料(含電子檔與老師簽名)並將附件報告平裝成壹冊交回系辦。(5-1)指導老師可依實際進度依序繳交 (a)專案執行規劃書(b) 需求規格報告書 (c) 系統設計報告書(d)系統測試報告，格式請參考下列或是自行參考軟體工程相關書籍格式並分階段分學期完成整個專案報告。**  **(6)學生請於2年級下學期期末考周前，完成分組與論文題目與摘要擬定並請指導教授簽名。** | |

**附件：參考格式-大綱 (本專案格式是以 1 主系統、3 個子系統、1 使用者操作介面為模型撰寫，完整參考報告請直 接跟系辦索取，指導教授可依指導專題特性彈性調整格式)**

**壹、專案執行規劃書**

**第1章 版本變更紀錄**

2022 年 11月 15日 第一版撰寫

2023 年 01 月 01日 第二版撰寫

2023 年 04 月 20日 第三版撰寫

**第2章 專案範圍**

**2.1 技術方法**

表1.技術方法表

|  |  |
| --- | --- |
| 模型名稱 | 參考論文 |
| AR | 郭合閔 [1] |
| ARMA |
| ARIMA |
| SARIMAX |
| Bayes Ridge Regression |
| Elastic net Regression |
| Polynomial Regression |
| RandomForest Regression | ellemcfarlane [2] |
| Support Vector Regression |

表1為本專題使用的模型名稱與對應的參考論文，本專題使用ARIMA系列傳統時間模型與5種機器學習模型對經濟項目進行預測。

**2.2 工作項目總表**

表2.工作項目總表

|  |  |
| --- | --- |
| 工作項目 | 工作內容 |
| 後端框架 | MariaDB |
| FLASK |
| REST API |
| BayesianRidge算法 | Sklearn |
| Polynomial Regression算法 |
| Elastic Net Regression算法 |
| Support Vector Regression |
| Random Forest Regression |
| Grid Search CV |
| Bayes Search CV | Skopt |

表2為本項目之工作項目總表，分為後端框架與算法。

後端架構:

本專題預計使用FLASK + MariaDB搭建資料庫系統並且透過REST API與主機進行連結，達到傳送數據之功能。

算法:

本專題於此學期加入5種機器學習算法進行預測及新增2種參數優化算法，進而探討每種模型之特性與優劣，參數優化方法能夠使模型的預測更加精準，本專題目前使用兩種參數優化算法，分別為:

1. Grid Search CV
2. Bayes Search CV

**2.3 專案製作動機與構想**

加密貨幣是基於區塊鏈技術的新型交易媒介，近年以比特幣及以太幣等主流加密貨幣，對全球商業經濟的影響力逐漸增長，受到政府部會及產業界的重視。分析金融經濟議題的典型方法為整合移動平均自回歸模型（Autoregressive Integrated Moving Average model）時間序列分析方法，然而，對於近年快速變化的複雜金融商業環境，時間序列分析方法的預測能力仍有提昇空間。本專題將探討3種時間序列分析方法、3種機器學習方法，並整合建構為一經濟時空分析平台，進行油價、匯率和股價等3種經濟指標對於加密貨幣的趨勢預測效能分析。本專題成果將提供一種經濟時空分析的解決方案，能夠進一步理解經濟指標和加密貨幣的影響及交互關係，並探討經濟指標在時間與空間上的不同效果，以供政府部會及產業擬定政策參考。

* 1. **使用情境**

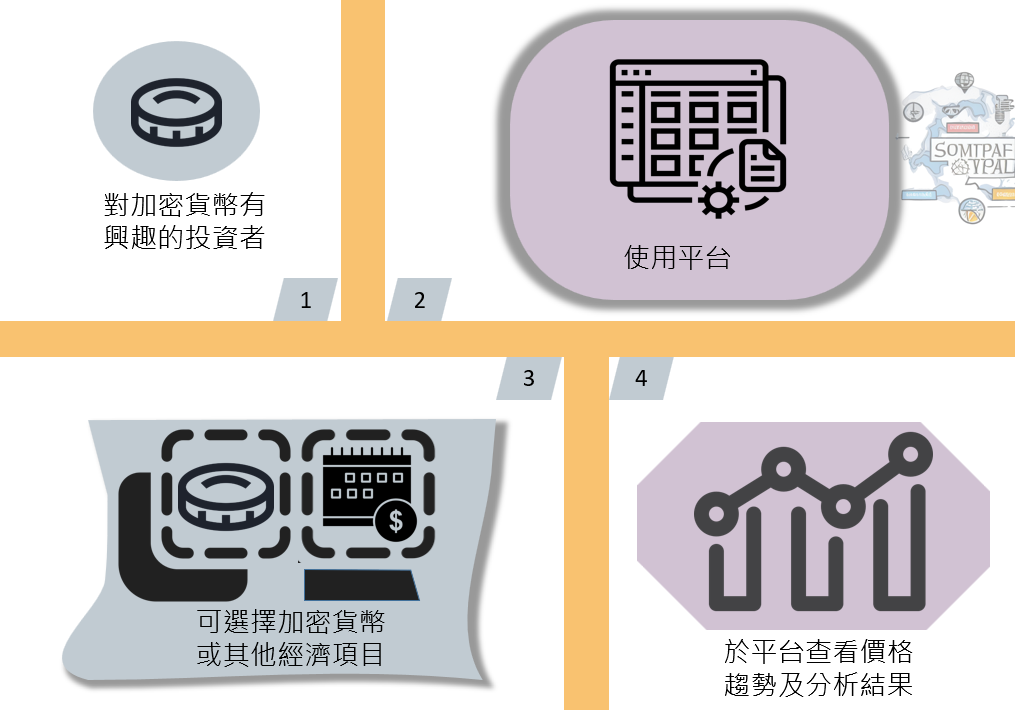
****

圖1. 系統功能圖

圖1為本專使用情境圖，其內容敘述為:對加密貨幣有興趣的投資者可以使用本平台價格趨勢圖與分析圖表，也可以自由選擇加密貨幣或經濟項目，並得到價格預測結果。

* 1. **資料集**

表3 資料集資訊表

|  |  |  |  |
| --- | --- | --- | --- |
|  | 日期 | 筆數 | 來源 |
| Bitcoin | 2013/04/29~2023/03/26 | 3619 | CoinMarketCap |
| Brent | 1987/05/20~2023/01/23 | 9056 | GitHub |
| WTI | 1986/01/02~2023/01/23 | 9337 | GitHub |
| Dubai | 2014/04/28~2022/06/03 | 2038 | GitHub |
| Gold | 2014/01/01~2022/08/05 | 2227 | Kaggle |

表3 為本專題之資料集資訊表，列出5種經濟項目的日期、筆數與來源。

**2.6 演算法**

1. Bayesian Ridge Regression

使用貝葉斯公式將觀測數據與先驗概率結合起來計算后驗概率分佈，從而得出最終的模型參數。與傳統的嶺回歸模型相比，Bayesian Ridge Regression模型在建模過程中引入了額外的超參數來控制模型的複雜度。這些超參數可以通過交叉驗證等技術進行調整，從而提高模型的預測性能。

1. Polynomial Regression

多項式回歸使用多項式函數（如二次、三次、四次等）來建立自變數和因變數之間的關係。多項式回歸模型可以通過最小二乘法等方法來估計模型參數，從而擬合出一個曲線，使得該曲線能夠最好地擬合數據。在實際應用中，多項式回歸模型通常需要通過交叉驗證等技術來確定多項式的次數，以避免過擬合或欠擬合的問題。

1. Elastic Net Regression

Elastic Net是一種線性回歸模型，它是L1正則化和L2正則化的組合。L1正則化可以讓模型參數中的某些特徵值稀疏化，而L2正則化可以讓模型參數更加平滑，從而可以有效地防止過擬合。相對於L1正則化和L2正則化分別使用的Lasso和Ridge Regression，Elastic Net的優點在於能夠選擇較多的相關特徵，並且在高維數據中不容易過擬合，同時也能保持模型參數的稀疏性。

1. Random Forest Regression

隨機森林是一種基於決策樹的集成學習方法，它通過組合多個決策樹來提高模型的準確性和穩定性。隨機森林中的每個決策樹都是在隨機子集上訓練的，這樣可以避免模型對特定的數據集過擬合。

1. Support Vector Machine

SVR基於支援向量機（SVM）演算法。與傳統的線性回歸模型不同，SVR通過最小化預測誤差和模型複雜度之間的權衡來尋找最優解。SVR使用一個核函數將輸入映射到一個高維空間中，使得在新的空間中的線性函數能夠更好地適應數據。

**2.7研究結果**

表4. GridSearch結果表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grid Search之結果 | | | | | |
|  | Bayesian Ridge Regression | Elastic Net | Polynomial Regression | SVR | Random Forest |
| RMSE | 9.91 | 8677.68 | 33.27 | 599.37 | 525.62 |
| MAPE | 0.20 | 20.73 | 0.07 | 2.13 | 1.14 |
| SMAPE | 0.20 | 23.28 | 0.07 | 2.10 | 1.14 |
| MAE | 8.64 | 7444.22 | 21.75 | 519.24 | 338.45 |
| R2 | 0.99 | 0.61 | 0.99 | 0.99 | 0.99 |
| Times | 12.15 | 0.04 | 1073.49 | 145.89 | 231.48 |

表5 Bayes Search結果表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Bayes Search之結果 | | | | | |
|  | Bayesian Ridge Regression | Elastic Net | Polynomial Regression | SVR | Random Forest |
| RMSE | 9.91 | 8476.12 | 457.13 | 599.37 | 505.11 |
| MAPE | 0.20 | 20.16 | 0.74 | 2.13 | 1.01 |
| SMAPE | 0.20 | 22.57 | 0.74 | 2.10 | 1.01 |
| MAE | 8.64 | 7256.87 | 225.37 | 519.24 | 311.95 |
| R2 | 0.99 | 0.63 | 0.99 | 0.99 | 0.99 |
| Times | 63.37 | 45.38 | 273.68 | 145.05 | 402.47 |

表4與表5為本專題機器學習算法之Grid Search結果表，本專題採用5種機器學習方法:

1. Bayesian Ridge Regression

Bayesian Ridge Regression是一種基於貝葉斯統計的線性回歸模型，Bayesian Ridge Regression與傳統線性回歸的不同之處在於Bayesian Ridge Regression引用了先驗概率分配，對模型進行建模，從而提高模型預測的精確度。

貝葉斯嶺回歸可用以下公式來表達:

1. Elastic Net

彈性網路(Elastic Net)為廣泛用於回歸分析的機器學習模型，它是Ridge Regression與 Lasso Regression之間的一種折衷方法。Elastic Net同時應用了L1及L2的正則化，它能夠應對高為數據下變量相關的情況，並提供較佳的性能。

Elastic Net可表示為:

1. Polynomial Regression

多項式回歸(Polynomial Regression)為線性回歸的拓展形式，它添加了多項式項來擬合非線性數據，多項式回歸的特點就是能夠解決非線性問題。

多項式回歸可以用以下式子來表達:

1. Support Vector Regression

支持向量回歸(Support Vector Regression)是一種監督式學習方法，與支持向量機(SVM)相似，但Support Vector Regression(SVR)用來解決回歸問題，SVR的目標是找到一個函數，此函數能夠最好的擬合給定的訓練數據集，並且在新數據集上有者很好的泛化能力，SVR的目標與SVM不同，SVR的目標是使得預測值及實際植之間的誤差能夠最小化。

支持向量回歸能夠以以下式子來表示:

Subject to :

1. Random Forest Regression

隨機森林(Random Forest)是一種集成機器學習算法，此方法廣泛的運用在分類及回歸問題上。隨機森林的基本運行思路是:創造多個決策樹，並以此來進行訓練及預測，每顆決策樹都是獨立訓練的，訓練時會使用不同的數據集，此方式可以降低過擬合的風險並提高模型的廣泛性，當要進行模型預測時，此模型會將各個決策樹的解果進行蒐集並總結，最終的採納結果也就是各個決策樹投票的結果。

對於分類問題，Random Forest使用基尼指數作為分裂質量的標準:

對於回歸問題，使用平均方差作為衡量分裂質量的標準:

與兩種參數調整工具:

1. Grid Search CV

網格搜尋(Grid Search)為一種常見的參數調整工具，在Grid Search中，使用者可以將需要挑整的超參數範圍都設定好，此工具就能夠幫你窮舉範圍內的每一種組合，最終得到結果最好的那一組參數。

1. Bayes Search CV

貝葉斯搜索(Bayesian Search)是一種參數調整工具，它通過先驗分布及後驗分布的貝葉斯公式，根據已有的觀測數據來調整最佳參數，與Grid Search相比，Bayesian Search不需要遍歷整個參數範圍，因此它能夠更高效的搜索參數組合。

來建構模型，並且採用五種指標:

1. RMSE
2. MAPE
3. SMAPE
4. MAE
5. R2

進行模型優劣之判斷，本專題將判斷的結果繪製成表格如表4與表5所示，研究結果表明Bayesian Ridge Regression 與 Elastic Net 在Grid Search CV的調整下於五種模型當中取得優勢，兩者的結果差距很小，Bayesian Ridge Regression與Elastic Net 為本專題在Grid Search CV基礎下的最佳模型。

**2.8 工作項目內容描述.**

表6. 工作項目內容描述表

|  |  |
| --- | --- |
| 項目 | 說明 |
| 平台前端設計 | 使用HTML, CSS, JavaScript 搭建前端 |
| 平台後端設計 | FLASK與MariaDB組成後端 |
| 機器學習模型訓練評估 | 使用5種機器學習模型訓練及評估 |
| 時間序列模型效能分析 | 使用ARIMA系列方法訓練及評估 |
| 整合平台模型API | 前端透過AJAX與FLASK與後端整合 |

表6為工作項目內容描述表分為五項:

1. 平台前端設計:
2. 平台後端設計: 使用XAMPP與PHP語言架設網站伺服器。
3. 機器學習模型訓練評估:使用五種機器學習模型Bayesian Ridge Regression、Elastic Net、Polynomial Regression、Support Vector Reg ression、Random Forest Regression。
4. 機器學習模型效能分析:使用五種評估指標(RMSE、MAPE、SMAPE、MAE、R2)評估執行結果。
5. 整合平台模型API：

**2.8.1預定查核點說明**

表7.預定查核表

|  |  |  |
| --- | --- | --- |
| 查核點 | 預定時間 | 查核點概述 |
| 1 | 2023/04/21 | 整理期末KPI |
| 2 | 2023/04/25 | 平台前端設計 |
| 3 | 2023/04/30 | 平台後端設計 |
| 4 | 2023/05/04 | 機器學習模型訓練評估 |
| 5 | 2023/05/10 | 機器學習模型效能分析 |
| 6 | 2023/05/18 | 整合平台模型API |

表6為本專題預定查核表。

查核點1為整理期末KPI。

查核點2為平台前端設計: 使用HTML, CSS, JavaScript 規劃及建構平台頁面。

查核點3為平台後端設計: 使用XAMPP與Python語言架設網站伺服器。

查核點4為機器學習模型訓練評估: 使用五種機器學習模型並提供三種經濟項目有Bitcoin, Global Crude Oil, Gold Price來預測選擇的經濟項目價格。

查核點5為機器學習模型效能分析: 使用五種評估指標有RMSE、MAPE、MAE、R2、SMAPE評估執行結果。

查核點6為整合平台模型API：

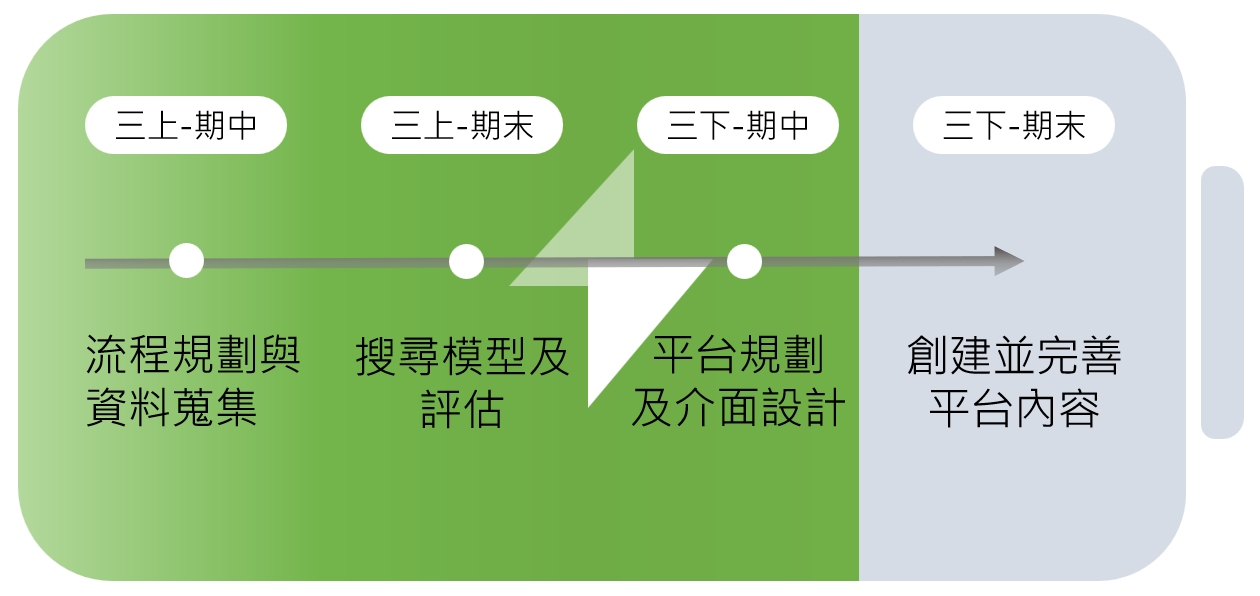
****

圖2.實際達成情形圖

圖2為本專題實際達成情形圖，在三上時完成流程規劃、資料蒐集、搜尋模型及評估，本次完成平台規劃與介面設計。

以上KPI都有符合時程規劃流程。

**第3章 資源**

**3.1 人力及工作分派**

表8. 人力及工作分派表

|  |  |
| --- | --- |
| 期末KPI | 任務指派 |
| 平台前端設計 | 鴻銘 |
| 平台後端設計 | 鴻銘 |
| 機器學習模型訓練評估 | 鴻銘 |
| 時間序列模型效能分析 | 鴻銘 |
| 整合平台模型API | 鴻銘 |

**文獻探討**

**[1-24]**

[1] 郭合閔, "以 Rapidminer 建構 ARIMA 時間序列預測與使用 B-Band 分析之投資效益比較-以加權股價指數為例," 碩士, 機電工程系, 國立高雄科技大學, 高雄市, 2022.

[2] ellemcfarlane, "Random Forest BTC," 2019.

[3] D. P. Ishan Kapur , Sajag Prakash, "BITCOIN PREDICTION USING MACHINE LEARNING," 2018.

[4] H. Jelicić, E. Phelps, and R. M. Lerner, "Use of missing data methods in longitudinal studies: the persistence of bad practices in developmental psychology," (in eng), *Dev Psychol,* vol. 45, no. 4, pp. 1195-9, Jul 2009.

[5] mertceyhan, "bitcoin-market-android," 2022.

[6] Patchett, "CryptoBuddy," 2019.

[7] SpiralDevelopment, "CryptoTracker," 2020.

[8] 方家齊, "自動機器學習是否提升油價可預測性研究-以西德油價爲例," 碩士, 經濟學系, 逢甲大學, 台中市, 2023.

[9] 白經睿, "比特幣價格與臺灣比特幣概念股股價的關聯," 碩士, 經濟學系經濟與財務碩士班, 淡江大學, 新北市, 2019.

[10] 任淑怡, "台灣景氣循環與國際原油價格-共整合及共特徵分析," 碩士, 經濟學研究所, 輔仁大學, 新北市, 2001.

[11] 沈彥彤, "頁岩油價格對於國際原油價格、黃金、美元報酬率之影響," 碩士, 財務金融學系碩士班, 淡江大學, 新北市, 2016.

[12] 林宜儒, "以股價和油價預測VIX指數," 碩士, 財務金融學系研究所, 國立嘉義大學, 嘉義市, 2020.

[13] 徐嘉臨, "原油價格、石油類股、太陽能類股動態關係之研究," 碩士, 企業管理學系碩士班, 開南管理學院, 桃園縣, 2006.

[14] 翁英鐘, "航空油價預測之研究," 碩士, 運籌管理學系, 國防大學, 桃園縣, 2016.

[15] 張庭瑜, "利用雲端機器學習預測原油價格及其應用於北美頁岩油產業," 碩士, 管理學院工業工程與管理學程, 國立陽明交通大學, 新竹市, 2022.

[16] 郭譽孚, "「日據時期」或「日治時期」."

[17] 陳淑玲, "石油價格與黃金價格衝擊對台灣加權股價指數期、現貨的影響," 碩士, 合作經濟學系, 國立臺北大學, 新北市, 2005.

[18] 陳淑華, "黃金價格與股價指數、石油價格及波動率指數關係之研究," 碩士, 風險管理與保險研究所, 逢甲大學, 台中市, 2011.

[19] 彭國政, "以分量迴歸探討Covid-19疫情前後，比特幣價格對台灣匯率與台灣比特幣概念股的影響," 碩士, 企業管理學系, 中華大學, 新竹市, 2022.

[20] 甯若蓁, "運用向量自我迴歸(VAR)探討波羅的海綜合指數與貨量、船噸、航油價格及GDP成長率之關聯性," 碩士, 國際財務金融碩士在職專班(IEMBA), 國立臺北大學, 新北市, 2022.

[21] 程美齡, "原油價格預測單一變數模型之比較," 碩士, 航運管理學系, 國立臺灣海洋大學, 基隆市, 2022.

[22] 黃姿穎, "油價、金價、匯率與國際股市之關聯性研究," 碩士, 財務金融學系, 義守大學, 高雄市, 2009.

[23] 劉銘騏, "運用機器學習方法分析比特幣交易行為與貨幣匯率之交互關係," 碩士, 資訊管理研究所, 國立交通大學, 2016.

[24] 藍崇瑜, "油價對經濟成長影響之探討—考慮國家特性差異的追蹤資料分量迴歸," 碩士, 國際企業學研究所, 國立臺灣大學, 台北市, 2022.