

BÀI TOÁN XOAY QUANH GDP VÀ HÀNH ĐỘNG

Trong một thế giới đầy biến động, nơi nền kinh tế toàn cầu liên tục chịu ảnh hưởng bởi các cú sốc bất định – từ đại dịch, khủng hoảng tài chính, xung đột địa chính trị cho tới biến đổi khí hậu – câu hỏi không chỉ còn là “GDP hiện tại bao nhiêu?”, mà trở thành: “Chúng ta có thể nhìn thấy trước điều gì?”

Các lý thuyết kinh tế học truyền thống đã trang bị cho chúng ta vô vàn công cụ phân tích tăng trưởng: từ mô hình ARIMA, VAR đến những chỉ số như CPI, tỷ lệ thất nghiệp, hay yield curve. Tuy nhiên, trong kỷ nguyên mà dữ liệu chảy về từng giờ, các mối quan hệ kinh tế có thể được nhìn nhận phi tuyến, phi cổ điển và thay đổi theo thời gian, thì khả năng dự báo động chính xác không thể chỉ dựa vào những mô hình tuyến tính cổ điển.

Đây là lúc Machine Learning và Deep Learning bước vào – không thay thế kinh tế học, mà mở rộng khả năng của kinh tế học, giúp chúng ta không chỉ mô hình hóa tương lai, mà còn hành động trước tương lai. Trong bài viết này, cùng SciData Vietnam tổng hợp 3 bài toán kinh điển liên quan đến GDP – từ dự báo tăng trưởng, phân loại chu kỳ – trạng thái kinh tế, cho đến nhận diện điểm ngoặt chu kỳ – kèm theo các kỹ thuật phân tích từ truyền thống đến hiện đại (ARIMA, VAR, LSTM, XGBoost,...), gợi ý cho các bạn đang nghiên cứu kinh tế học ứng dụng, chính sách công, tài chính vĩ mô hay đang triển khai sản phẩm data-driven trong bối cảnh kinh tế.

Điều quan trọng không chỉ là xây mô hình giỏi, mà là: Nếu biết được GDP sẽ chứng lại, bạn sẽ làm gì? Nếu phát hiện sớm khả năng suy thoái, bạn sẽ chuẩn bị ra sao? Hãy cùng đi sâu vào dữ liệu, tìm hiểu thuật toán – và quan trọng nhất: hành động dựa trên những gì chúng ta học được từ dự báo.

I. BA BÀI TOÁN VỀ GDP

1. Dự báo tăng trưởng GDP (Real GDP Forecasting)

Dự báo GDP theo quý hoặc năm (YoY) là một trong những bài toán kinh điển nhất trong kinh tế học thực nghiệm. Thông qua việc ước lượng gần chính xác đến chính xác mức tăng trưởng GDP trong tương lai, Chính phủ và các nhà chính sách có cơ sở để hoạch định chính sách tiền tệ, tài khóa, và lập kế hoạch đầu tư cho khu vực công và tư.

Mô tả: Dự báo giá trị của GDP thực (sau điều chỉnh lạm phát) ở mốc thời gian tiếp theo, theo % thay đổi năm trên năm (YoY).

Phân loại: Hồi quy (Regression)

Một số mô hình và nền tảng:

Mô hình	Nền tảng hoạt động	Ghi chú
ARIMA/ SARIMA	R (<code>forecast</code> , <code>tsseries</code>) Python (<code>statsmodels</code>)	Mô hình tuyến tính, hiệu quả cho chuỗi đơn biến
ARX/ SARIMAX	Python (<code>statsmodels.tsa.statespace.sarimax</code>)	Bổ sung biến giải thích như CPI, unemployment
Ridge/Lasso	Python (<code>scikit-learn</code>)	Kiểm soát overfitting và chọn biến phù hợp
XGBoost/ LightGBM	Python (<code>xgboost</code> , <code>lightgbm</code>)	Dự báo phi tuyến hiệu quả, mạnh với dữ liệu đa chiều
LSTM/ GRU	Python (<code>TensorFlow</code> , <code>Keras</code> , <code>PyTorch</code>)	Dự báo chuỗi dài hạn, nắm bắt được phụ thuộc thời gian phức tạp

Đầu ra kỳ vọng & Ý nghĩa:

- **Giá trị GDP dự báo (YoY %):** Cung cấp góc nhìn chính sách, phân tích chu kỳ, dự phòng rủi ro vĩ mô.
- **Giải thích được yếu tố ảnh hưởng:** Biến nào dự báo tốt hơn? Tác động tăng/giảm ra sao?

Case study:

- Giannone et al. (2008): Áp dụng mô hình Factor + Ridge Regression dự báo GDP của Mỹ thời gian thực (“nowcasting”), độ sai số thấp hơn nhiều so với mô hình AR.
- Foroni et al. (2020): So sánh giữa ARIMA, LSTM và các mô hình ML trong dự báo GDP châu Âu; LSTM vượt trội trong giai đoạn có cú sốc COVID-19.

Lưu ý khi phân tích bài toán:

- Không nên dùng split ngẫu nhiên (dùng TimeSeriesSplit hoặc Walk-forward CV).
- Chuỗi GDP cần được kiểm tra tính dừng và dừng hóa nếu dùng ARIMA.
- Với LSTM: cần chuẩn hoá, xử lý giá trị khuyết thiếu, kiểm soát hiện tượng overfitting.
- Cần đặc biệt lưu ý outlier (ví dụ: COVID, chiến tranh) vì gây nhiễu mô hình ML mạnh và giảm ý nghĩa dự báo của đầu ra mô hình.

Ý nghĩa thực tiễn:

- Hỗ trợ dự báo tăng trưởng quốc gia cho ngân sách, quy hoạch công.
- Làm cơ sở cho các quỹ đầu tư quốc gia/địa phương và các tác nhân khác trong nền kinh tế ra quyết định phân bổ vốn và điều chỉnh các yếu tố kinh tế vĩ mô.

2. Phân loại trạng thái kinh tế – Nhận biết suy thoái (Recession Detection / Classification)

Xác định nền kinh tế hiện đang ở trong thời kỳ nào là vấn đề sống còn với chính sách vĩ mô và đầu tư tài chính. Trong khi nhiều mô hình chỉ tập trung vào nhận diện suy thoái (recession), thực tế nền kinh tế còn trải qua giai đoạn tăng trưởng nóng (overheating), dễ dẫn đến bong bóng tài sản, lạm phát cao và các rủi ro tài chính tiềm ẩn. Nhận diện đúng trạng thái kinh tế hiện tại – từ suy thoái đến tăng trưởng quá mức – là cơ sở cho điều hành chính sách tài khoá và tiền tệ linh hoạt, đảm bảo ổn định nền kinh tế vĩ mô.

Mô tả bài toán: Dự đoán nền kinh tế tại thời điểm t đang thuộc vào một trong ba trạng thái sau:

- Recession – Suy thoái
- Neutral / Expansion – Tăng trưởng ổn định
- Overheating – Tăng trưởng nóng, tiềm ẩn rủi ro lạm phát

Lưu ý: Một số nghiên cứu có thể thay từ “Overheating” bằng “(Economic) Boom”, nhưng “Overheating” là thông dụng hơn trong cảnh báo rủi ro chính sách.

Phân loại bài toán: Phân loại đa lớp (multi-class classification)

Một số mô hình và nền tảng:

Mô hình	Nền tảng	Ghi chú
Multinomial Logistic Regression	R Python (statsmodels , sklearn)	Dễ giải thích, cơ sở tốt
Random Forest / Extra Trees	Python (sklearn)	Mạnh với dữ liệu không tuyến tính và nhiều chiều
Gradient Boosting (XGBoost / LightGBM)	Python R	Hiệu quả cao với dữ liệu macro có tương tác phi tuyến
Multi-Class SVM	Python (sklearn)	Mạnh trong việc phân chia rõ biên ranh giới
LSTM + Softmax	Python (Keras , PyTorch)	Dùng cho chuỗi thời gian, khai thác cấu trúc động ẩn

Đầu vào bài toán: Các chỉ số kinh tế vĩ mô, bao gồm:

- CPI & core inflation
- Đường cong lợi suất (yield spread: 10y – 2y, hoặc 10y – Fed Funds Rate)
- Tỷ lệ thất nghiệp
- ISM PMI / IPI / Business Confidence Index
- M2 (trong chính sách tiền tệ) / tốc độ tăng tín dụng
- Price-to-Earnings ratio (nếu thêm yếu tố tài chính)
- Output Gap (nếu có khả năng tính toán)

Đầu ra kỳ vọng & Ý nghĩa:

- **Nhân phân loại:** Xác định chính xác trạng thái kinh tế (Recession/Neutral/Overheating)
- Chỉ báo định hướng chính sách:
 - Nếu recession → kích thích tài khóa / tiền tệ.
 - Nếu overheating → thắt chặt tiền tệ, kiểm soát tín dụng.
 - Tín hiệu thị trường tài chính: Tái phân bổ danh mục đầu tư, phòng ngừa rủi ro chu kỳ.

Case study:

- Koop et al. (2009): Áp dụng mô hình Markov-Switching + Random Forest để phân loại 3 pha: suy thoái – ổn định – bùng nổ (recession – normal – boom) cho dữ liệu GDP và lạm phát của OECD → Cho thấy xác suất “boom” tăng vọt trong các giai đoạn trước khủng hoảng tài chính 2008 tại Mỹ, Ireland và Tây Ban Nha.
- Bruneau, de Bandt & Savard (2020): Dùng mô hình Gradient Boosting phát hiện nguy cơ tăng trưởng nóng (credit overextension) tại khu vực đồng Euro (Eurozone) thông qua phân tích các chỉ số tín dụng, lạm phát và đầu tư.

Lưu ý khi phân tích bài toán:

(1) Class imbalance nghiêm trọng hơn bài toán nhị phân:

- Thời gian nền kinh tế “overheating” rất ngắn so với “neutral”.

(2) Dán nhãn (labeling) phải rõ ràng và khách quan:

- Dữ liệu recession có thể dùng từ NBER.
- Overheating nên dựa vào ngưỡng như: $CPI > 4\%$, $unemployment < NAIRU$, $output\ gap > 2\%$ GDP tiềm năng (theo IMF / OECD).

(3) Biến đầu vào nên mang tính chu kỳ mạnh:

- Chọn các biến có tính chất leading, không phải chỉ phản ứng chậm.

(4) Mô hình phi tuyến mạnh cần giải thích đi kèm (SHAP / Permutation Importance):

- Nhất là trong ứng dụng cho hoạch định chính sách công.

(5) Luôn kiểm tra mô hình với dữ liệu thời kỳ “không bình thường” (COVID, khủng hoảng tài chính, xung đột quân sự, xung đột địa chính trị):

- Dự báo có thể bị sai lệch nặng nếu không có dữ liệu tương đương để huấn luyện.

3. Dự báo điểm ngoặt chu kỳ GDP (Turning Point Prediction)

Không chỉ dự đoán GDP, điều quan trọng hơn là xác định khi nào nền kinh tế sẽ đổi chiều – bước vào giai đoạn phục hồi hay suy thoái. Đây là thông tin chiến lược để chính phủ và doanh nghiệp ra quyết định đúng lúc.

Mô tả bài toán: Phát hiện và dự đoán thời điểm chuyển pha từ tăng trưởng sang suy thoái (hoặc ngược lại).

Phân loại bài toán: Phân loại tuần tự hoặc phát hiện dị thường theo chuỗi (sequential classification / anomaly detection).

Một số mô hình và nền tảng:

Mô hình	Nền tảng hoạt động	Ghi chú
Markov Switching Model (MSM)	R (MSwM) / Python (statsmodels)	Mô hình phổ biến phát hiện chuyển pha
Hidden Markov Model (HMM)	Python (hmmlearn)	Xác suất chuyển pha động
LSTM / BiLSTM	Python (Keras , PyTorch)	Mạnh với dữ liệu thời gian dài
Hybrid (ARIMA + ML)	R / Python	ARIMA xác định xu hướng, ML dự báo residual để bắt turning point

Đầu ra kỳ vọng và Ý nghĩa:

- **Xác định đúng điểm đảo chiều:** Kích hoạt cảnh báo sớm, giảm rủi ro cho cả chính sách và thị trường tài chính.
- **Biểu đồ xác suất chuyển pha:** Cho phép dự đoán điểm ngoặt sớm hơn thống kê chính thức (thường công bố chậm).

Case study:

- Chauvet & Potter (2013): Sử dụng MSM kết hợp LSTM để xác định điểm ngoặt GDP trong khủng hoảng tài chính 2008 tại Mỹ – LSTM phát hiện sớm hơn công bố NBER tới 2 quý.
- Clements & Galvão (2008): HMM nhận diện điểm chuyển pha GDP tại Anh hiệu quả hơn mô hình tuyến tính.

Lưu ý khi phân tích bài toán:

- MSM/HMM: Cần giả định số pha (regime), thường là 2 hoặc 3.
- LSTM: Phát hiện sớm tốt nhưng cần explainability nếu ứng dụng trong chính sách.
- Cần đối chiếu kết quả với dữ liệu bên ngoài (business cycle chronologies) để xác minh độ tin cậy.

II. LỜI KẾT

Dù có mục tiêu khác nhau, cả 3 bài toán xoay quanh GDP đều có điểm chung: tác động sâu rộng đến chính sách công, tài chính, và doanh nghiệp. Khi áp dụng, cần chọn mô hình phù hợp với đặc thù dữ liệu và mục đích sử dụng – ưu tiên tính chính xác nhưng không bỏ qua khả năng diễn giải và ứng dụng thực tế.

III. TÀI LIỆU THAM KHẢO

Ang, A., Piazzesi, M., & Wei, M. (2006). What does the yield curve tell us about GDP growth? *Journal of Econometrics*, 131(1–2), 359–403. <https://doi.org/10.1016/j.jeconom.2005.01.032>

Baffigi, A., Golinelli, R., & Parigi, G. (2002). Leading indicators and the Italian business cycle. *Economic Modelling*, 19(3), 611–629. [https://doi.org/10.1016/S0264-9993\(01\)00068-4](https://doi.org/10.1016/S0264-9993(01)00068-4)

Ban, T., Konishi, T., & Naito, T. (2013). Financial time series forecasting by K-nearest neighbor regression: A case study in Japanese stock market. *International Journal of Computer Applications*, 74(5), 1–6. <https://doi.org/10.5120/12921-9952>

Bellotti, T., Huang, Y., & Crook, J. (2021). Forecasting economic recessions using hybrid models. *Expert Systems with Applications*, 178, 114976. <https://doi.org/10.1016/j.eswa.2021.114976>

Bruneau, C., de Bandt, O., & Savard, K. (2020). Predicting financial overheating and credit cycles using machine learning. *Banque de France Working Paper No. 763*. <https://doi.org/10.2139/ssrn.3696049>

Chauvet, M., & Potter, S. (2013). Forecasting output. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (Vol. 2, pp. 141–194). Elsevier. <https://doi.org/10.1016/B978-0-444-53683-9.00004-5>

Diebold, F. X., Rudebusch, G. D., & Aruoba, S. B. (2006). The macroeconomy and the yield curve: A dynamic latent factor approach. *Journal of Econometrics*, 131(1–2), 309–338. <https://doi.org/10.1016/j.jeconom.2005.01.009>

Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *The Journal of Finance*, 46(2), 555–576. <https://doi.org/10.1111/j.1540-6261.1991.tb02629.x>

Foroni, C., Marcellino, M., & Schumacher, C. (2020). Forecasting economic activity with mixed frequency models. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (Vol. 2B, pp. 285–339). Elsevier. <https://doi.org/10.1016/bs.hesmac.2020.05.003>

Koop, G. (2013). Forecasting in macroeconomics. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (Vol. 2A, pp. 523–657). Elsevier. <https://doi.org/10.1016/B978-0-444-53683-9.00012-4>

Koop, G., Leon-Gonzalez, R., & Strachan, R. W. (2009). Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric Reviews*, 28(1–3), 171–197. <https://doi.org/10.1080/07474930802388074>

Rodríguez-Vargas, J. A. (2020). Forecasting GDP using Random Forests: Evidence from the euro area. *Empirical Economics Letters*, 19(8), 827–842. [Available through institutional access or repositories]

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, 5(4), 595–620. <https://doi.org/10.1214/aos/1176343886>