

DỰ BÁO LẠM PHÁT BẰNG MÔ HÌNH TRUYỀN THỐNG VÀ MÁY HỌC

Khi nói đến lạm phát, mọi người đều nghĩ ngay đến chỉ số giá tiêu dùng (CPI) hay sự tăng liên tục và phổ biến của mức giá hàng hóa và dịch vụ trong một nền kinh tế trong cùng một khoảng thời gian xác định. Lạm phát thường có tác động tiêu cực đến phát triển kinh tế - xã hội. Tuy nhiên, nếu nền kinh tế đó thích nghi tốt với sự thay đổi của lạm phát bằng các giải pháp thích hợp thì có thể phòng ngừa và hạn chế tối đa thiệt hại cho nền kinh tế cũng như có thể khai thác triệt để mặt tích cực của lạm phát trong một số trường hợp. Vì thế, **dự báo lạm phát** là một trong những nhiệm vụ quan trọng hàng đầu, không chỉ đối với các nhà hoạch định chính sách mà còn cho cả các doanh nghiệp, nhà đầu tư và người dân.

Trong kỷ nguyên mà dữ liệu rất đa dạng và phong phú, các mối quan hệ có ảnh hưởng đến lạm phát 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.

Do đó, áp dụng Machine Learning và Deep Learning sẽ giúp chúng ta đánh giá hiệu quả hơn và có những hướng đi phù hợp hơn trong tương lai. Trong bài viết này, cùng Scidata Vietnam tổng hợp 7 bài toán xoay quanh đến lạm phát và ứng dụng các mô hình từ truyền thống đến hiện đại để hỗ trợ ra quyết định hiệu quả hơn trong bối cảnh kinh tế hiện nay và gợi ý cho các bạn đang nghiên cứu kinh tế học ứng dụng, hay các vấn đề liên quan tới lĩnh vực kinh tế.

Như đã đề cập, dự báo lạm phát không chỉ giới hạn ở việc đưa ra một con số đơn lẻ. Tùy thuộc vào mục tiêu cụ thể của người sử dụng, bài toán lạm phát có thể được tiếp cận dưới nhiều góc độ khác nhau, dẫn đến các dạng bài toán khác nhau. Việc phân loại rõ ràng các dạng bài toán giúp chúng ta lựa chọn phương pháp và mô hình phù hợp nhất để dễ dàng tìm ra các câu trả lời cho: Đối tượng là gì? Tại sao? Và bằng cách nào? Dưới đây là 7 dạng bài toán phổ biến, hãy cùng đi sâu vào tìm hiểu thuật toán, dữ liệu và cách ứng dụng vào thực tế.

I. CÁC DẠNG BÀI TOÁN VỀ LẠM PHÁT

1. Dự báo giá trị CPI tương lai (*Future CPI Values Forecasting*)

Dự báo chỉ số giá tiêu dùng (CPI) là hoạt động quan trọng đối với không chỉ chính phủ mà cả các doanh nghiệp trong việc lập kế hoạch hoạt động. Kết quả dự báo càng chính xác thì việc lập kế hoạch càng khả thi.

- Mô tả:** Ước tính giá trị cụ thể của chỉ số CPI (hoặc tỉ lệ lạm phát %) tại một thời điểm xác định trong tương lai.
- Phân loại:** Mô hình hồi quy (Regression)
- Một số mô hình nền tảng:**

Mô hình	Nền tảng hoạt động	Ghi chú
ARIMA / SARIMA	R (forecast), Python (statsmodels)	Mô hình chuỗi thời gian đơn biến, xử lý tốt các cấu trúc động tuyến tính. SARIMA thêm yếu tố thời gian.

ARIMAX / SARIMA	Python (statsmodels)	Mở rộng ARIMA, cho phép đưa biến giải thích ngoài (giá dầu, tỷ giá,...).
VAR ? VECM	R, Python (statsmodels, arch)	Mô hình hóa quan hệ động giữa nhiều biến vĩ mô (CPI, lãi suất, tỷ giá,...).
Linear Regression	R, Python (sklearn, statsmodels)	Cơ bản, dễ giải thích.
Ridge / Lasso	Python (sklearn)	Hồi quy tuyến tính có tiêu chuẩn, giúp xử lý nhiều biến, giảm overfitting.
XGBoost / LightGBM	Python, R	Mô hình phi tuyến mạnh mẽ, hiệu quả với dữ liệu đa chiều có tương tác.
LSTM / GRU	Python (TensorFlow, PyTorch)	Mạng nơ - ron hồi quy, mạnh với chuỗi thời gian dài, nắm bắt phụ thuộc phức tạp.
Hybrid Models	R, Python	Kết hợp mô hình truyền thống (như ARIMA) với ML (như XGBoost, LSTM).

d. Đầu ra kỳ vọng & ý nghĩa:

- ↳ Giá trị CPI hoặc tỷ lệ lạm phát dự báo: Cung cấp con số mục tiêu cho các kế hoạch và dự báo vĩ mô / vi mô.
- ↳ Khoảng tin cậy dự báo: Đánh giá mức độ bất định của dự báo.

e. Case study:

- Giannone et al. (2008): Dù tập trung vào GDP, phương pháp Factor models cùng với Ridge Regression của họ là nền tảng cho nhiều mô hình “nowcasting” (dự báo tức thì) các chỉ số vĩ mô, bao gồm cả CPI, tận dụng dữ liệu tần suất cao.
- Smith et al. (2018): Nghiên cứu so sánh hiệu quả của ARIMA, VAR, Random Forest và LSTM trong dự báo CPI ở nhiều quốc gia, chỉ ra rằng các mô hình ML thường vượt trội trong các giai đoạn biến động lớn.

f. Lưu ý khi phân tích bài toán:

- ↳ Cần kiểm tra tính dừng (stationarity) của chuỗi CPI và các biến đầu vào nếu sử dụng mô hình truyền thống (ARIMA, VAR).
- ↳ Lựa chọn các biến đầu vào (features) có tính dẫn dắt hoặc đồng biến với CPI.

- ↳ Xử lý độ trễ (lags) của các biến đầu vào (ví dụ: giá dầu, tỷ giá thường có tác động trễ đến CPI).
- ↳ Sử dụng phương pháp kiểm định chéo phù hợp cho chuỗi thời gian (TimeSeriesSplit, Walk-Forward).
- ↳ Cân trọng với các giá trị ngoại lai (outliers) hoặc cú sốc lớn (khủng hoảng) có thể ảnh hưởng đến hiệu suất mô hình, đặc biệt là ML.

g. Ý nghĩa thực tiễn:

- ↳ Căn cứ cho việc xây dựng chính sách tiền tệ (điều chỉnh lãi suất) và chính sách tài khóa (ngân sách, đầu tư công).
- ↳ Hỗ trợ doanh nghiệp trong việc lập kế hoạch giá bán, dự báo chi phí đầu vào, và quản lý tồn kho.
- ↳ Cung cấp thông tin cho nhà đầu tư để điều chỉnh danh mục, đặc biệt với các tài sản nhạy cảm với lạm phát (trái phiếu, hàng hóa).

2. Phân loại mức độ lạm phát (Inflation Level Classification)

Lạm phát đi từ đơn giản đến phức tạp sẽ bao gồm 3 mức độ chính, được phân loại dựa theo tỷ lệ phần trăm của lạm phát

a. Mô tả:

- *Lạm phát tự nhiên:* tương ứng với tỷ lệ lạm phát từ 0 – <10%. Nền kinh tế lúc này hoạt động bình thường xảy ra ít rủi ro và đời sống ổn định.
- *Lạm phát phi mã:* Là mức độ lạm phát xảy ra với tình trạng giá cả tăng nhanh, tỷ lệ từ 10 – < 1.000%. Loại này rất dễ gây biến động nền kinh tế.
- *Siêu lạm phát:* là tình trạng lạm phát tăng nhanh, khó kiểm soát với tốc độ chóng mặt, tỷ lệ trên 1.000%. Siêu lạm phát để lại hậu quả to lớn, khó lòng có thể khắc phục. Tuy nhiên, trường hợp siêu lạm phát rất hiếm khi xảy ra.

b. **Phân loại:** Phân loại (Classification) – thường là Phân loại đa lớp (Multi-class Classification).

c. Một số mô hình nền tảng:

Mô hình	Nền tảng hoạt động	Ghi chú
Logistic Regression	R, Python (statsmodels, sklearn)	Dễ giải thích, nền tảng cho bài toán phân loại. Multinomial cho đa lớp.
Random Forest / Extra Trees	Python (sklearn)	Mạnh mẽ, xử lý được tương tác phi tuyến, ít nhạy cảm với nhiễu.
Gradient Boosting (XGBoost / LightGBM)	R, Python	Hiệu quả cao, thường cho kết quả tốt với dữ liệu có cấu trúc phức tạp.

Support Vector Machine (SVM)	Python (sklearn)	Hiệu quả trong việc tìm biên phân cách tối ưu.
Neural Networks (MLP, CNN, LSTM + Softmax)	Python (Keras, PyTorch)	Nắm bắt quan hệ phức tạp, đặc biệt LSTM/CNN cho dữ liệu chuỗi/hình ảnh.

d. Đầu ra kỳ vọng & ý nghĩa:

- ↳ Nhận phân loại: Cho biết nền kinh tế đang ở nhóm lạm phát nào.
- ↳ Xác suất thuộc từng nhóm: Đánh giá mức độ chắc chắn của phân loại.
- ↳ Chỉ báo cho chính sách: Giúp Ngân hàng Trung ương xác định cần thắt chặt hay nới lỏng tiền tệ; Chính phủ điều chỉnh chi tiêu.
- ↳ Chỉ báo cho đầu tư: Định hướng phân bổ tài sản vào các loại hình phù hợp với từng giai đoạn lạm phát.

e. Case study:

- ↳ Medeiros et al. (2021): Áp dụng các mô hình ML như Random Forest và Boosting để phân loại các "chế độ" (regimes) kinh tế, bao gồm các chế độ lạm phát khác nhau, cho thấy ML có thể nhận diện các ranh giới phức tạp hơn mô hình tuyến tính.
- ↳ Central Bank Research Papers: Nhiều ngân hàng trung ương sử dụng các mô hình phân loại (có thể là Markov Switching) để nhận diện các pha của lạm phát và đưa ra chính sách phản ứng.

f. Lưu ý khi phân tích bài toán:

- ↳ Việc định nghĩa và dán nhãn (labeling) các nhóm lạm phát cần rõ ràng, có thể dựa trên ngưỡng lịch sử, mục tiêu chính sách, hoặc phân tích chuyên gia.
- ↳ Bài toán thường gặp vấn đề mất cân bằng lớp (class imbalance), do các giai đoạn lạm phát cực đoan (rất cao hoặc giảm phát) thường ngắn hơn giai đoạn lạm phát ổn định. Cần sử dụng kỹ thuật xử lý mất cân bằng dữ liệu.
- ↳ Lựa chọn các biến đầu vào có ý nghĩa kinh tế và có khả năng phân biệt các nhóm lạm phát.

g. Ý nghĩa thực tiễn:

- ↳ Hỗ trợ Ngân hàng Trung ương xác định "chế độ" lạm phát hiện tại để điều chỉnh công cụ chính sách (lãi suất, dự trữ bắt buộc).
- ↳ Giúp nhà đầu tư hiểu bối cảnh lạm phát để chọn chiến lược bảo vệ giá trị tài sản (ví dụ: chuyển sang hàng hóa, bất động sản khi lạm phát cao).
- ↳ Doanh nghiệp có thể dự đoán rủi ro lạm phát để điều chỉnh chiến lược giá và quản lý chi phí.

3. Dự báo xác suất lạm phát vượt ngưỡng (Forecasting Probability of Exceeding Threshold)

Bài toán này đặc biệt quan trọng đối với các Ngân hàng Trung ương có mục tiêu lạm phát, hoặc các tổ chức/doanh nghiệp muốn đánh giá rủi ro lạm phát vượt qua một ngưỡng nhất định.

- a. **Mô tả:** Ước lượng xác suất (probability) mà lạm phát (thường đo bằng CPI % YoY) sẽ vượt qua một ngưỡng mục tiêu hoặc rủi ro cụ thể trong tương lai (ví dụ: xác suất CPI > 5% trong 12 tháng tới).
- b. **Phân loại:** Phân loại nhị phân (Binary Classification) hoặc Hồi quy (với đầu ra là phân phối xác suất).
- c. **Một số mô hình nền tảng:**

Mô hình	Nền tảng hoạt động	Ghi chú
Logistic Regression	R, Python (statsmodels, sklearn)	Cơ sở tốt cho phân loại nhị phân, đầu ra là xác suất.
Gradient Boosting (XGBoost / LightGBM)	R, Python	Mạnh mẽ, thường cho xác suất được hiệu chỉnh tốt.
Bayesian Models	R (rstanarm, brms), Python (PyMC3)	Cung cấp phân phối xác suất đầy đủ, tính toán sự bất định.
Quantile Regression	R, Python (statsmodels)	Dự báo các điểm phân vị (quantile) của phân phối lạm phát.

- d. **Đầu ra kỳ vọng & ý nghĩa:**
 - ↳ Giá trị xác suất (Probability score): Tỷ lệ phần trăm khả năng lạm phát vượt ngưỡng.
 - ↳ Ý nghĩa: Cung cấp tín hiệu cảnh báo sớm, giúp Ngân hàng Trung ương quyết định có nên hành động để kiểm soát lạm phát hay không, giúp các nhà đầu tư định lượng rủi ro "đuôi" (tail risk) của lạm phát cao.
- e. **Case study:**
 - ↳ Ngân hàng Dự trữ Liên bang Mỹ (FED) và các Ngân hàng Trung ương lớn thường công bố các biểu đồ "Dot Plot" hoặc phân phối xác suất lạm phát để truyền thông về kỳ vọng và rủi ro.
 - ↳ Các công ty quản lý tài sản sử dụng mô hình dự báo xác suất để xây dựng các sản phẩm phòng ngừa rủi ro lạm phát hoặc các chiến lược giao dịch phái sinh dựa trên xác suất lạm phát.
- f. **Lưu ý khi phân tích bài toán:**
 - ↳ Độ chính xác xác suất (Probability Calibration): Quan trọng hơn chỉ đúng/sai, mô hình cần đưa ra xác suất phản ánh đúng mức độ tin cậy.
 - ↳ Lựa chọn ngưỡng: Ngưỡng lạm phát cần được định nghĩa rõ ràng (ví dụ: 3%, 5%, mục tiêu của NHTW).
 - ↳ Đánh giá mô hình: Sử dụng các metrics phù hợp với bài toán phân loại nhị phân và quan tâm đến xác suất (ví dụ: AUC, Brier Score, Log Loss).

g. Ý nghĩa thực tiễn:

- ↳ Cơ sở để Ngân hàng Trung ương áp dụng các biện pháp tiền tệ sớm (pre-emptive tightening) hoặc truyền thông định hướng kỳ vọng thị trường.
- ↳ Hỗ trợ các tổ chức tài chính thiết kế các sản phẩm bảo hiểm hoặc phòng ngừa rủi ro lạm phát.

4. Phát hiện điểm ngoặt (Turning Point Prediction)

Khả năng nhận diện sớm khi lạm phát chuẩn bị đổi chiều xu hướng (từ tăng sang giảm hoặc ngược lại) là thông tin cực kỳ giá trị, cho phép các chủ thể kinh tế hành động kịp thời, tránh bỏ lỡ cơ hội hoặc giảm thiểu thiệt hại.

- a. Mô tả:** Phát hiện và dự đoán thời điểm mà tốc độ tăng lạm phát sẽ đạt đỉnh hoặc chạm đáy và bắt đầu chuyển hướng.
- b. Phân loại:** Phát hiện thay đổi cấu trúc (Change Point Detection), Phân loại tuần tự (Sequential Classification) hoặc Dự báo điểm cực trị (Extrema Forecasting).
- c. Một số mô hình nền tảng:**

Mô hình	Nền tảng hoạt động	Ghi chú
Markov Switching Model (MSM)	R (MSwM), Python (statsmodels)	Phổ biến trong phân tích chuyển pha kinh tế.
Hidden Markov Model (HMM)	Python (hmmlearn)	Mô hình xác suất cho chuỗi thời gian ẩn trạng thái.
Bayesian Change Point Methods	R, Python (ruptures, bayesian-changepoint)	Phát hiện điểm thay đổi cấu trúc dựa trên xác suất.
LSTM / BiLSTM	Python (Keras, PyTorch)	Mạnh mẽ trong nhận diện các mẫu phức tạp trong chuỗi thời gian.
Hybrid Models (ARIMA + ML)	R, Python	Kết hợp để bắt cả xu hướng tuyến tính và sai số phi tuyến.
Dynamic Factor Models with Regimes	R, Python	Sử dụng các nhân tố ẩn có trạng thái chuyển đổi.

d. Đầu ra kỳ vọng & ý nghĩa:

- ↳ Tín hiệu/cảnh báo về điểm ngoặt tiềm năng: Chỉ ra thời điểm lạm phát có khả năng đổi chiều.
- ↳ Xác suất đang ở trong giai đoạn chuyển đổi: Định lượng sự bất định xung quanh tín hiệu.

↳ Ý nghĩa: Giúp Ngân hàng Trung ương định thời điểm tuyên bố "đỉnh lạm phát" hoặc "đáy lạm phát" để định hướng chính sách và truyền thông, giúp doanh nghiệp/nhà đầu tư đưa ra quyết định chiến lược về giá, hàng tồn kho, hay phân bổ tài sản trước khi xu hướng lạm phát mới thiết lập.

e. Case study:

- ↳ Nhiều nghiên cứu sử dụng các mô hình HMM hoặc MSM để phân tích lịch sử lạm phát, xác định các giai đoạn lạm phát cao/thấp và thời điểm chuyển đổi giữa chúng.
- ↳ Các quỹ phòng hộ sử dụng các mô hình dự báo điểm ngoặt dựa trên Machine Learning để tối ưu hóa thời điểm ra/vào các vị thế liên quan đến lạm phát.

f. Lưu ý khi phân tích bài toán:

- ↳ Định nghĩa "Điểm ngoặt": Có thể dựa trên đỉnh/đáy cục bộ của chuỗi lạm phát đã được làm mượt (smoothed) hoặc các tiêu chí khác.
- ↳ Độ nhạy cảm của mô hình: Các mô hình phát hiện điểm ngoặt có thể nhạy cảm với dữ liệu nhiễu, cần tiền xử lý cẩn thận.
- ↳ Đánh đổi giữa Cảnh báo sớm và Độ chính xác: Mô hình cảnh báo quá sớm có thể có tỷ lệ cảnh báo sai (false positive) cao.

g. Ý nghĩa thực tiễn:

- ↳ Tối ưu hóa thời điểm thực hiện các biện pháp chính sách tiền tệ/tài khóa.
- ↳ Giúp doanh nghiệp điều chỉnh chiến lược mua sắm, sản xuất, định giá kịp thời.
- ↳ Cung cấp tín hiệu giao dịch chiến lược cho các thị trường tài chính.

5. Nowcasting CPI (Dự báo tức thời CPI)

Trong khi dữ liệu CPI chính thức thường được công bố với độ trễ (ví dụ: CPI tháng 3 được công bố vào giữa tháng 4), Nowcasting (dự báo tức thời) nhằm mục đích ước lượng giá trị CPI của kỳ hiện tại (ví dụ: ước lượng CPI tháng 4 ngay trong tháng 4) dựa trên các dữ liệu có tần suất cao hơn và được công bố sớm hơn.

- a. Mô tả:** Ước lượng giá trị của CPI cho giai đoạn hiện tại (ví dụ: tháng, quý) trước khi dữ liệu chính thức được công bố, sử dụng thông tin từ các chỉ báo kinh tế được cập nhật thường xuyên hơn.
- b. Phân loại:** Hồi quy (Regression), Kết hợp dữ liệu đa tần suất (Mixed-Frequency Data).
- c. Một số mô hình nền tảng:**

Mô hình	Nền tảng hoạt động	Ghi chú
Mixed Data Sampling (MIDAS) Models	R (midasr), Python (statsmodels)	Chuyên dụng xử lý dữ liệu có tần suất khác nhau.
Dynamic Factor Models (with mixed freq)	R, Python	Rút trích nhân tố chung từ dữ liệu tần suất cao để dự báo biến tần suất thấp.
State-Space Models	R, Python (statsmodels)	Cho phép tích hợp thông

		tin từ các nguồn dữ liệu khác nhau theo thời gian.
XGBoost / LightGBM	Python (xgboost, lightgbm)	Áp dụng trên các đặc trưng được tổng hợp từ dữ liệu tần suất cao.
Time-Series Transformers / Deep Learning	Python (TensorFlow, Keras, PyTorch)	Các mô hình học sâu có thể xử lý trực tiếp dữ liệu đa tần suất.

d. Đầu ra kỳ vọng & ý nghĩa:

- ↳ Ước lượng giá trị CPI tức thời: Cung cấp bức tranh kinh tế hiện tại "gần như thời gian thực".
- ↳ Ý nghĩa: Cực kỳ quan trọng cho các nhà hoạch định chính sách và giao dịch viên tài chính cần phản ứng nhanh chóng với diễn biến kinh tế. Giúp đánh giá tác động của các sự kiện vừa xảy ra lên lạm phát ngay lập tức

e. Case study:

- ↳ Các Ngân hàng Trung ương (ví dụ: ECB, FED thông qua các phòng nghiên cứu) và các ngân hàng đầu tư thường xuyên thực hiện nowcasting CPI/GDP bằng các mô hình phức tạp để hỗ trợ ra quyết định.
- ↳ Các bài nghiên cứu áp dụng mô hình MIDAS hoặc Factor Models cho nowcasting các biến vĩ mô.

f. Lưu ý khi phân tích bài toán:

- ↳ Tính sẵn có và độ tin cậy của dữ liệu tần suất cao: Nguồn dữ liệu cần ổn định và có liên quan thực sự đến CPI.
- ↳ Xử lý dữ liệu khuyết thiếu và cập nhật: Dữ liệu tần suất cao thường có độ trễ hoặc bị sửa đổi sau này.
- ↳ Mô hình hóa mối quan hệ giữa các tần suất: Đây là thách thức cốt lõi của nowcasting.

g. Ý nghĩa thực tiễn:

- ↳ Cung cấp thông tin kịp thời cho các cuộc họp chính sách.
- ↳ Hỗ trợ các quyết định giao dịch tài chính ngắn hạn.
- ↳ Đánh giá tác động kinh tế của các cú sốc bất ngờ.

6. Phân rã đóng góp vào CPI (CPI Contribution Decomposition)

Bài toán này không trực tiếp dự báo giá trị CPI, mà tập trung vào việc phân tích *nguyên nhân* đằng sau sự thay đổi của nó. Hiểu rõ thành phần nào (ví dụ: năng lượng, lương thực, dịch vụ) hoặc yếu tố nào (ví dụ: giá dầu, tỷ giá, tiền lương) đang đóng góp bao nhiêu vào tổng thay đổi CPI giúp đưa ra các giải pháp chính sách hoặc chiến lược kinh doanh phù hợp.

- a. **Mô tả:** Xác định mức độ đóng góp của từng thành phần trong rổ hàng hóa/dịch vụ của CPI (ví dụ: thực phẩm, nhà ở, năng lượng...) hoặc mức độ đóng góp của từng biến giải thích (ví dụ: giá dầu, tỷ giá, M2) vào sự thay đổi hoặc mức độ của CPI.
- b. **Phân loại:** Giải thích mô hình (Model Interpretation), Phân tích đóng góp (Contribution Analysis).
- c. **Một số mô hình nền tảng:**

Mô hình	Nền tảng hoạt động	Ghi chú
Linear Models (Regression)	R, Python (statsmodels, sklearn)	Hệ số hồi quy (coefficients) trực tiếp cho biết đóng góp (với dữ liệu chuẩn hóa).
Decision Tree	Python (sklearn), R	Có thể theo dõi đường dẫn quyết định để hiểu đóng góp của biến.
SHAP (SHapley Additive exPlanations)	Python (shap)	Kỹ thuật giải thích cho "hộp đen", phân bổ đóng góp của mỗi biến cho mỗi dự báo.
LIME (Local Interpretable Model-agnostic Explanations)	Python (lime)	Giải thích dự báo cục bộ, cho biết biến nào quan trọng cho một điểm dữ liệu cụ thể.
Structural Decomposition	Phân tích kinh tế học	Phân tích dựa trên cấu trúc kinh tế để phân rã thay đổi tổng thể.
VAR Impulse Response Functions	R (vars), Python (statsmodels)	Phân tích tác động của một cú sốc lên các biến khác theo thời gian.

- d. **Đầu ra kỳ vọng & ý nghĩa:**
 - ↳ Biểu đồ/bảng phân rã đóng góp: Minh họa rõ ràng nguồn gốc của lạm phát.
 - ↳ Giá trị đóng góp của từng yếu tố: Số liệu định lượng về mức độ ảnh hưởng.

↳ Ý nghĩa: Giúp nhận diện "điểm nóng" gây lạm phát (ví dụ: năng lượng tăng vọt), hỗ trợ thiết kế các chính sách đối phó (ví dụ: giảm thuế môi trường với nhiên liệu), hoặc giúp doanh nghiệp/nhà đầu tư tập trung phân tích vào các yếu tố chi phối chính.

e. Case study:

- ↳ Các báo cáo CPI chính thức thường có phần phân tích đóng góp của các nhóm hàng chính.
- ↳ Các nhà kinh tế sử dụng SHAP/LIME để giải thích tại sao mô hình ML của họ dự báo lạm phát cao hoặc thấp trong một giai đoạn cụ thể.
- ↳ Phân tích Impulse Response từ mô hình VAR để hiểu tác động của cú sốc tỷ giá lên CPI tổng thể và các thành phần.

f. Lưu ý khi phân tích bài toán:

- ↳ Mô hình cơ sở: Phương pháp giải thích phụ thuộc vào mô hình dự báo đang sử dụng.
- ↳ Tương tác biến: Các mô hình phức tạp (ML) có thể có tương tác biến, cần kỹ thuật giải thích phù hợp (như SHAP) thay vì chỉ xem xét hệ số đơn giản.
- ↳ Độ chi tiết dữ liệu: Cần dữ liệu chi tiết theo thành phần để thực hiện phân rã theo nhóm hàng.

g. Ý nghĩa thực tiễn:

- ↳ Hỗ trợ đưa ra các chính sách vĩ mô có mục tiêu (ví dụ: kiềm chế lạm phát do chi phí đẩy).
- ↳ Doanh nghiệp hiểu rõ động lực chi phí của mình.
- ↳ Nhà đầu tư phân tích các rủi ro theo ngành/yếu tố cụ thể.

7. Dự báo kỳ vọng lạm phát (Inflation Expectation Forecasting)

Kỳ vọng lạm phát (Inflation Expectation) là dự đoán của công chúng, doanh nghiệp và thị trường tài chính về mức lạm phát trong tương lai. Kỳ vọng này có vai trò quan trọng vì nó có thể trở thành yếu tố tự thực hiện (self-fulfilling prophecy) và ảnh hưởng trực tiếp đến hành vi kinh tế (người lao động yêu cầu tăng lương, doanh nghiệp tăng giá bán, nhà đầu tư yêu cầu lãi suất cao hơn). Do đó, dự báo kỳ vọng lạm phát cũng là một bài toán riêng biệt và quan trọng.

- a. Mô tả:** Ước lượng mức lạm phát mà các tác nhân kinh tế kỳ vọng sẽ xảy ra trong tương lai (ví dụ: kỳ vọng lạm phát 1 năm tới của người tiêu dùng, kỳ vọng 5-10 năm tới của thị trường tài chính).
- b. Phân loại:** Hồi quy (Regression - nếu đầu ra là số), Phân tích cảm xúc (Sentiment Analysis - nếu từ văn bản), Kết hợp NLP và Regression.
- c. Một số mô hình nền tảng:**

Mô hình	Nền tảng hoạt động	Ghi chú
---------	--------------------	---------

Regression Models (trên dữ liệu khảo sát/thị trường)	R, Python (statsmodels, sklearn)	Sử dụng các biến vĩ mô để dự báo giá trị kỳ vọng.
Sentiment Analysis (NLP)	Python (nltk, spacy, transformers)	Phân tích thái độ/cảm xúc về lạm phát từ văn bản (báo cáo, tin tức, mạng xã hội).
BERT / LSTM + Regression/Sentiment	Python (TensorFlow, PyTorch)	Các mô hình học sâu cho xử lý ngôn ngữ tự nhiên, kết hợp với dự báo số.
Yield Curve Analysis	R, Python	Trích xuất kỳ vọng lạm phát từ đường cong lợi suất của trái phiếu (ví dụ: TIPS).

d. Đầu ra kỳ vọng & ý nghĩa:

- ↳ Giá trị kỳ vọng lạm phát (ví dụ: 2.5% 1 năm tới).
- ↳ Chỉ số cảm xúc (sentiment score) liên quan đến lạm phát từ dữ liệu phi cấu trúc.
- ↳ Ý nghĩa: Giúp Ngân hàng Trung ương đánh giá mức độ neo giữ (anchoring) của kỳ vọng lạm phát, làm cơ sở cho việc truyền thông và định hướng chính sách tiền tệ. Giúp doanh nghiệp hiểu tâm lý thị trường/người tiêu dùng.

e. Case study:

- ↳ Các Ngân hàng Trung ương theo dõi chặt chẽ các chỉ số kỳ vọng lạm phát từ khảo sát và thị trường tài chính.
- ↳ Nghiên cứu sử dụng NLP để phân tích phát biểu của quan chức FED hoặc ECB nhằm đo lường tác động đến kỳ vọng lạm phát.
- ↳ Sử dụng mô hình chuỗi thời gian để dự báo lợi suất trái phiếu TIPS và từ đó suy ra kỳ vọng lạm phát ngụ ý từ thị trường.

f. Lưu ý khi phân tích bài toán:

- ↳ Nguồn dữ liệu đa dạng: Cần kết hợp thông tin từ nhiều nguồn (khảo sát, thị trường, văn bản) để có cái nhìn toàn diện.
- ↳ Tính chủ quan của khảo sát: Kết quả khảo sát có thể bị ảnh hưởng bởi tâm lý nhất thời.

- ↳ Độ nhiễu trên thị trường: Kỳ vọng lạm phát từ thị trường tài chính có thể bị ảnh hưởng bởi các yếu tố khác ngoài lạm phát thuần túy (ví dụ: thanh khoản, rủi ro).
- ↳ Thách thức của NLP: Xử lý ngôn ngữ tự nhiên đòi hỏi kỹ thuật phức tạp để trích xuất thông tin liên quan lạm phát.

g. Ý nghĩa thực tiễn:

- ↳ Hỗ trợ Ngân hàng Trung ương trong việc truyền thông chính sách (forward guidance) và quản lý neo kỳ vọng.
- ↳ Doanh nghiệp đánh giá tâm lý khách hàng và người lao động liên quan đến giá cả và tiền lương.
- ↳ Nhà đầu tư phân tích rủi ro và cơ hội trên thị trường trái phiếu.

II. CÁC MÔ HÌNH TRUYỀN THỐNG DÙNG ĐỂ DỰ BÁO CPI

Các mô hình kinh tế lượng cổ điển đã là xương sống của dự báo lạm phát trong nhiều thập kỷ. Chúng dựa trên nền tảng lý thuyết kinh tế và phân tích chuỗi thời gian, cung cấp sự ổn định và khả năng diễn giải tốt.

- **ARIMA / SARIMA:** Phù hợp cho dự báo dựa trên lịch sử của chính chuỗi CPI, hiệu quả khi lạm phát có xu hướng và tính mùa vụ rõ ràng. Dễ triển khai và diễn giải các tham số. Tuy nhiên, chúng bỏ qua tác động của các biến kinh tế vĩ mô khác và nhạy cảm với các thay đổi cấu trúc lớn.
- **ARIMAX / SARIMAX:** Khắc phục nhược điểm của ARIMA bằng cách kết hợp các biến giải thích bên ngoài (exogenous variables) như giá dầu, tỷ giá, hoặc các chỉ số hoạt động kinh tế. Giúp mô hình hóa tốt hơn các kênh truyền dẫn lạm phát.
- **VAR / VECM:** Cho phép phân tích mối quan hệ động giữa nhiều biến vĩ mô cùng lúc, cung cấp một cái nhìn hệ thống. Đặc biệt hữu ích khi các biến có tác động qua lại phức tạp. VECM được dùng khi các biến có mối quan hệ cân bằng dài hạn.
- **Phillips Curve:** Mô hình hóa mối quan hệ giữa lạm phát và thị trường lao động (hoặc output gap). Vẫn là công cụ quan trọng để hiểu động lực lạm phát từ phía cầu kéo và chi phí tiền lương.

→ Ưu điểm chung của nhóm này là tính diễn giải (interpretability) cao, phù hợp với nền tảng lý thuyết kinh tế. Nhược điểm là khả năng nắm bắt các mối quan hệ phi tuyến hoặc sự thay đổi cấu trúc đột ngột có thể hạn chế.

III. ỨNG DỤNG MACHINE LEARNING VÀO DỰ BÁO CPI

Machine Learning mở ra khả năng xử lý lượng dữ liệu lớn và đa dạng hơn, phát hiện các mẫu hình phức tạp và phi tuyến mà mô hình truyền thống khó nắm bắt.

- **Linear Regression, Lasso, Ridge:** Là bước khởi đầu tốt cho các bài toán hồi quy. Lasso và Ridge giúp quản lý số lượng biến lớn và chống overfitting thông qua điều chuẩn (regularization).
- **Random Forest, XGBoost, LightGBM:** Các mô hình dựa trên cây (tree-based) cực kỳ mạnh mẽ với dữ liệu có tương tác phi tuyến. Chúng thường mang lại độ chính xác cao và có thể cung cấp thông tin về mức độ quan trọng của biến.

- **LSTM / GRU / Bi-LSTM:** Đặc biệt hiệu quả cho chuỗi thời gian dài. Khả năng "ghi nhớ" thông tin từ quá khứ giúp chúng nắm bắt tốt các xu hướng dài hạn và các phụ thuộc phức tạp theo thời gian, đặc biệt trong các bài toán phát hiện điểm ngoặt.
- **Hybrid Model (Kết hợp truyền thống + ML):** Tận dụng ưu điểm của cả hai phương pháp. Ví dụ: dùng ARIMA để dự báo xu hướng tuyến tính, dùng ML để dự báo phần sai số còn lại; hoặc sử dụng đầu ra mô hình kinh tế lượng làm đặc trưng (feature) cho mô hình ML. Phương pháp này thường cải thiện độ chính xác và giữ được một phần khả năng diễn giải.

Ưu điểm của nhóm này là khả năng xử lý dữ liệu phức tạp, tự động học hỏi các mẫu hình phi tuyến, và thường cho độ chính xác dự báo cao hơn trong môi trường nhiều biến động. Nhược điểm là tính "hộp đen" (black box) của nhiều mô hình, đòi hỏi kỹ thuật bổ trợ để giải thích.

IV. KỸ THUẬT NÂNG CAO ĐỂ TỐI ƯU MÔ HÌNH

Để xây dựng một hệ thống dự báo mạnh mẽ, cần kết hợp các kỹ thuật tối ưu hóa và đánh giá:

- **GridSearchCV, RandomSearchCV, Bayesian Optimization:** Các phương pháp dò tìm siêu tham số (hyperparameter tuning) để tìm cấu hình tốt nhất cho mô hình.
- **Cross-validation (TimeSeriesSplit):** Phương pháp kiểm định chéo phù hợp với chuỗi thời gian để đánh giá đúng hiệu suất dự báo ngoài mẫu (out-of-sample).
- **SHAP / LIME:** Kỹ thuật giải thích mô hình ML (Explainable AI - XAI), giúp hiểu biến nào đóng góp bao nhiêu vào mỗi dự báo cụ thể, giảm tính "hộp đen".
- **Ensemble & Stacking:** Kết hợp dự báo từ nhiều mô hình độc lập khác nhau để cải thiện độ mạnh mẽ và độ chính xác tổng thể.
- **Attention mechanism (Deep Learning):** Cho phép các mô hình Deep Learning tập trung vào các phần quan trọng nhất của dữ liệu quá khứ khi đưa ra dự báo, cải thiện khả năng nắm bắt các sự kiện quan trọng.

V. MỘT VÀI LƯU Ý KHI XÂY DỰNG MÔ HÌNH DỰ BÁO CPI

Thực tế triển khai mô hình dự báo lạm phát cần lưu ý nhiều khía cạnh kỹ thuật và dữ liệu:

- **Tần suất dữ liệu:** CPI là chỉ số hàng tháng, do đó dữ liệu đầu vào cũng nên được thu thập hoặc tổng hợp về tần suất tháng để phản ánh đúng động lực lạm phát.
- **Độ trễ biến đầu vào:** Cần xác định và xử lý độ trễ phù hợp cho các biến giải thích (ví dụ: giá dầu, tỷ giá) vì tác động của chúng lên CPI thường không tức thời.
- **Tiền xử lý dữ liệu:** Bao gồm xử lý giá trị khuyết thiếu (missing values), ngoại lai (outliers), và chuẩn hóa (scaling) dữ liệu để phù hợp với yêu cầu của mô hình.
- **Tối ưu hóa hàm mất mát (loss function):** Lựa chọn hàm mất mát phù hợp với mục tiêu bài toán (ví dụ: MAE/RMSE cho hồi quy, F1/AUC cho phân loại).

VI. LỜI KẾT

Không có một mô hình "đũa thần" duy nhất cho mọi bài toán dự báo lạm phát. Một chiến lược dự báo CPI hiện đại và hiệu quả đòi hỏi sự kết hợp hài hòa giữa hiểu biết sâu sắc về kinh tế học, nền tảng vững chắc về phân tích chuỗi thời gian truyền thống, và sức mạnh tính toán cùng khả năng học hỏi từ dữ liệu của Machine Learning.

Vì thế, để xây dựng một hệ thống dự báo lạm phát mạnh mẽ, linh hoạt và đáng tin cậy, chúng ta nên:

- Sử dụng **Hybrid Model** để tận dụng ưu điểm của cả mô hình truyền thống và ML
- Đầu tư vào **Feature Engineering** chất lượng cao, đặc biệt là việc lựa chọn và xử lý các biến có tính dự báo và độ trễ phù hợp.
- Áp dụng các kỹ thuật **Explainable AI** (SHAP, LIME) để tăng tính minh bạch và giải thích cho các mô hình ML, điều này đặc biệt quan trọng khi kết quả dự báo được dùng cho hoạch định chính sách hoặc ra quyết định chiến lược.

Nội dung Học Thuật: Như Ý

Thông tin liên hệ

➤ Email: scidatavietnam@gmail.com

➤ Fanpage: <https://www.facebook.com/SciDataVietnam>

TÀI LIỆU THAM KHẢO

[1] GS.TS Nguyễn Thanh Tuyền, & Đặng Hữu Thủy. (2014). *Luận văn Thạc sĩ Kinh tế: Lạm phát và ứng dụng mô hình Arima để dự báo lạm phát ở Việt Nam*.
<https://tailieu.vn/doc/luan-van-thac-si-kinh-te-lam-phat-va-ung-dung-mo-hinh-arima-de-du-bao-lam-phat-o-viet-nam-2883201.html>

[2] Huyền N. T. T., & Sơn N. T. (2022). *DỰ BÁO CHỈ SỐ GIÁ TIÊU DÙNG TẠI VIỆT NAM BẰNG MÔ HÌNH ARIMA*.

[3] Juha Junntila. (2001). *Structural breaks, ARIMA model and Finnish inflation forecasts*.
<https://www.sciencedirect.com/science/article/abs/pii/S0169207000000807>

[4] Nextech. Các phương pháp tinh chỉnh siêu tham số trong mô hình trí tuệ nhân tạo AI. *Các phương pháp tinh chỉnh siêu tham số trong mô hình trí tuệ nhân tạo AI*.
<https://nextech.vn/cac-phuong-phap-tinh-chinh-sieu-tham-so-trong-mo-hinh-tri-tue-nhan-tao-ai/>

[5] Nhóm tác giả Trường Đại Học Sư Phạm Kỹ Thuật Thành Phố Hồ Chí Minh. (2014). *Nghiên cứu mô hình ARIMA và ANN trong dự báo lạm phát tại Việt Nam*.

<https://vn-document.net/document/du-bao-lam-phat-viet-nam-arima-ann/8934747759>

[6] Saranya Gangadhara Moorthy. (2024, Tháng Năm). *Grid Search based Optimum Feature Selection by Tuning hyperparameters for Heart Disease Diagnosis in Machine learning*.

https://www.researchgate.net/publication/371760914_Grid_Search_based_Optimum_Feature_Selection_by_Tuning_hyperparameters_for_Heart_Disease_Diagnosis_in_Machine_learning

[7] ThS. Nguyễn Thị Hiền, TS. Lê Mai Trang, Phạm Long Vũ, Phạm Văn Duy Hoàng, Lê Quang Huy & Trường Đại học Thương mại. (2023, Tháng Năm 15). *Ứng dụng mô hình ARIMA để dự báo lạm phát của Việt Nam và một số khuyến nghị*.

<https://tapchinganhang.gov.vn/ung-dung-mo-hinh-arima-de-du-bao-lam-phat-cua-viet-nam-va-mot-so-khuyen-nghi-12311.html>

[8] Trần Đài. (2021). Lạm phát là gì? Phân loại các mức độ lạm phát và ảnh hưởng tới kinh tế Việt Nam. *Lạm phát là gì? Phân loại các mức độ lạm phát và ảnh hưởng tới kinh tế Việt Nam*.

<https://cafedautu.vn/lam-phat-la-gi-phan-loai-cac-muc-do-lam-phat-va-anh-huong-toi-kinh-te-viet-nam-d2196.html>

[9] TS. ĐINH BẢO NGỌC. (2015). *Tóm tắt Luận văn Ứng dụng mô hình Arima trong dự báo lạm phát Việt Nam*.

<https://doan.edu.vn/do-an/tom-tat-luan-van-ung-dung-mo-hinh-arima-trong-du-bao-lam-phat-viet-nam-40055/>