Project Evolution & Methodology Report

EA Macro Deceleration Early Warning

By: Paulina Mamiaga

---

1. Initial Project Idea

Original Objective

The initial objective was to predict **GDP growth (t+1)** for Emerging Market (EM) economies using macroeconomic indicators from the IMF World Economic Outlook database.

The original formulation was a **regression problem**:

> Predict real GDP growth one year ahead using macro fundamentals.

2. Dataset Selection and Initial Work (Emerging Markets Phase)

Data Source

IMF World Economic Outlook (WEO)

- Annual macroeconomic data
- Long panel structure
- Wide cross-country coverage
- Multiple macro indicators

Initial Steps Performed

1. Filtered countries classified as Emerging Markets.
2. Selected relevant macroeconomic indicators.

3. Pivoted dataset from long to wide format.
4. Cleaned missing values.
5. Built lag structure (features at t → target at t+1).
6. Constructed GDP growth t+1 target.
7. Ran univariate analysis and correlation matrix.
8. Trained baseline OLS regression.

Initial Findings

The regression results showed:

- Low out-of-sample R^2
- Weak predictive power
- High multicollinearity
- Instability in coefficients
- Small effective sample after cleaning (~500 observations)

Additionally:

- Emerging Markets data had heterogeneous coverage.
- Some countries had structural breaks.
- Data gaps reduced panel balance.

3. Strategic Reassessment

Given:

- Weak regression performance
- Data instability

- Heterogeneous macro structures
- Reduced effective observations
- Limited time before deadline

A strategic reassessment was conducted.

This decision was taken autonomously to ensure project feasibility within the available time.

4. Pivot in Strategy

Major Changes Implemented

1. Region Change

From: Emerging Markets
To: European economies (Euro Area + selected peers)

Rationale:

- More homogeneous macroeconomic structure
- Higher data coverage consistency
- More synchronized business cycles
- Reduced structural noise

2. Target Redefinition

Original:
Predict GDP growth level (regression)

New:
Predict growth deceleration (classification)

New target:

GDP_accel_tplus1 = GDP(t+1) − GDP(t)
Decel_flag_tplus1 = 1 if acceleration < 0

This transforms the problem into:

Early Warning of Growth Deceleration

3. Model Type Change

Original: Regression
New: Binary Classification

Reason:

- Growth acceleration is more stationary than growth level.
- Early warning models are more interpretable.
- Classification showed stronger signal.
- Out-of-time AUC significantly higher than regression R^2.

5. Methodology – New Framework

Region

Selected European economies:

DEU, FRA, ITA, ESP, NLD, BEL, AUT, FIN, IRL, PRT, GRC
Plus additional smaller economies for sample expansion.

Time range:
1980–2023

Effective ML dataset:
737 observations

Feature Selection Process

Steps:

1. Coverage analysis (top variables by availability).
2. Correlation matrix analysis.
3. Multicollinearity screening.

4. Removal of leakage variables.
5. Final feature selection (9 macro fundamentals).

Final features include:

- Unemployment
- Inflation
- Trade growth
- Fiscal balance
- Current account
- GDP growth
- Output gap
- Labor productivity

Data Preparation Steps

- Filtered actual observations (removed forecasts).
- Pivoted long → wide.
- Created lag structure (t → t+1).
- Constructed acceleration target.
- Built binary deceleration flag.
- Imputed missing values where necessary.
- Removed duplicates.
- Verified no leakage.
- Ensured balanced class distribution (~48%).

Final ML-ready dataset:
737 rows × 13 columns
No missing values
No duplicates

6. Key Findings from EDA

1. Deceleration clusters around known crises:
    - Early 1990s
    - 2008–2009
    - 2020
2. GDP acceleration shows cyclical mean-reverting behavior.
3. Macroeconomic variables co-move during crises.
4. The binary target is balanced.
5. The heatmaps confirm synchronized downturns.

This validated the economic relevance of the target.

7. Comparison: Initial vs Revised Strategy

Dimension	Emerging Markets Regression	EA Classification
Stability	Low	High
Homogeneity	Low	High
Sample size	~500	737
Signal strength	Weak	Stronger
Interpretability	Moderate	High
Crisis alignment	Noisy	Clear

The revised approach provides:

- Stronger predictive signal
- More robust economic interpretation
- Cleaner panel structure
- Better alignment with early-warning literature

8. Current Phase of the Project

Completed:

- Dataset selection
- Data cleaning
- Feature engineering
- Target construction
- EDA
- Documentation
- ML-ready dataset export
- Project restructuring
- Repository organization

Delivered to ML teammate:

- Clean CSV
- Data dictionary
- ML handoff document

- Target definition
- Feature list
- Time split recommendation

9. What Remains to Be Done

Modeling Phase (Next Steps)

- Train baseline Logistic Regression
- Perform time-based validation
- Compute AUC, F1, Recall
- Tune hyperparameters
- Compare with ensemble models
- Select best model

Advanced Steps

- Probability threshold optimization
- Feature importance analysis
- Cross-validation
- Robustness checks

Deployment

- Streamlit app or demo interface
- Model serialization

- Demo scenario preparation

Presentation Preparation

- Slides
- Model comparison visualization
- Executive summary
- Q&A preparation
- Rehearsal

10. Project Status

Current stage:

Data engineering phase complete.
ML modeling phase beginning.

Project is:

Structurally sound
Economically coherent
Technically organized
ML-ready

11. Strategic Reflection

The pivot from Emerging Markets regression to European classification was driven by:

- Empirical evidence
- Data quality constraints
- Model performance evaluation
- Time management considerations

The revised framework is:

- ☑ More coherent
- ☑ More defendable
- ☑ More aligned with macroeconomic early-warning systems