INTERNATIONAL MONETARY FUND

Leveraging Non-traditional Data for Macroeconomic Nowcasting: The Case of Morocco

Dina Hamed

WP/26/108

IMF Working Papers **describe research in progress by the author(s) and are published to elicit comments and to encourage debate.**
The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

2026
JUN

WORKING PAPER

© 2026 International Monetary Fund WP/26/108

IMF Working Paper
Middle East and Central Asia Department

Leveraging Non-traditional Data for Macroeconomic Nowcasting: The Case of Morocco

Prepared by Dina Hamed*

Authorized for distribution by Laura Jaramillo Mayor
June 2026

IMF Working Papers **describe research in progress by the author(s) and are published to elicit comments and to encourage debate.** The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

ABSTRACT: Making informed policy decisions is contingent upon the availability of reliable and timely data. The use of non-traditional data has been shown to be a powerful tool for enabling policymakers to conduct robust nowcasting—the practice of estimating the current period's economic indicator(s), ahead of official releases, using a wide range of macroeconomic and high-frequency data. This paper showcases how different types of non-traditional data, such as indices extracted from satellite imagery, Google Trends, and flight tracking information, can be leveraged to complement official statistics and monitor economic activity, and how these timely signals can be incorporated into nowcasting models to provide early estimates of key macroeconomic variables in Morocco. The approach is applied to agricultural gross value added, tourism revenues, and the unemployment rate. The results demonstrate that non-traditional data substantially improves nowcasting models by enhancing predictive accuracy and enabling the rapid generation of nowcast estimates prior to the release of official data.

RECOMMENDED CITATION: Hamed, D. (2026). Leveraging Non-traditional Data for Macroeconomic Nowcasting: The Case of Morocco. Working Paper. International Monetary Fund WP/26/108

JEL Classification Numbers:	C22, C52, C53, C55, E37
Keywords:	Nowcasting; Macroeconomic Forecasting; Non-traditional data; Satellite Imagery; Google Trends; Tourism Revenues; Agriculture GVA; Unemployment Rate; Machine learning; Morocco
Author's E-Mail Address:	dhamed@imf.org

* The author gratefully acknowledges Laura Jaramillo, Marco Marini, and Niall O'Hanlon for their continued support and valuable guidance, and notes with appreciation the support and feedback of Nadia Mounir, Kassia Antoine, Marzie Taheri Sanjani, Hannah Claire Brown, the IMF MCD Morocco team, Alessandra Sozzi, Iyke Maduako, the IMF STA BDC team, Allen Boddie, Anthony Silungwe, Hany Abdel-Latif, and Nikolay Danov; and the editing support of Sofia Cerna Rubinstein. The author would also like to thank the Moroccan authorities for their cooperation.

WORKING PAPERS

Leveraging Non-traditional Data for Macroeconomic Nowcasting: The Case of Morocco

Prepared by Dina Hamed

Contents

Glossary **4**

Executive Summary **5**

Introduction **6**

2. Literature Review **8**

3. Non-Traditional Data **10**
3.1 Agriculture Value Added 10
3.2 Tourism Revenues 15
3.3. Unemployment Rate 18

4. Methodology **19**
4.1. Data Transformation and Processing 19
4.1.1. Agriculture Value Added 19
4.1.2. Tourism Revenues 19
4.1.3. Unemployment Rate 19
4.2. Nowcasting Framework 20
4.2.1 Models 20
4.2.2. Hyperparameter Tuning and Model Evaluation 22
4.2.3. Robustness Checks 23

5. Results **23**
5.1. Agriculture Value Added 24
5.2. Tourism Revenues 26
5.3 Unemployment Rate 27

6. Conclusions **30**

References **31**

FIGURES

1. Agricultural Stress Index (ASI) Heatmap by District 12
2. Precipitation in Morocco, January 2025 vs. January 2026 12
3. Correlation between NDVI and Cereal Production 13
4. Nighttime Lights, 2015 vs. 2025 13
5. Correlations between Satellite Indicators and Agriculture Value Added 14
6. Tourism Arrivals and Passenger Capacity (FlightRadar24) 15
7. Breakdown of Passenger Capacity across Major Airports during the 2025 Africa Cup of Nations 16
8. Tourism Revenues and Selected Google Trends Queries: Correlations 17
9. Tourism Revenues and Selected Google Trends Queries: Time-Series Comparison 17
10. Unemployment Rate and Selected Google Trends Queries: Correlations 18
11. Unemployment Rate and Selected Google Trends Queries: Time-Series Comparison 18

12. Agriculture GVA Q1/26 Nowcasts by Vintage 24
13. Agriculture GVA Nowcasting Plots 25
14. Shapley Decomposition of Q1/26 Nowcast 26
15. Tourism Revenues Nowcasting Plots 27
16. Tourism Revenues Nowcasting ARIMAX Model Coefficients 27
17. Unemployment Rate Nowcasting Plots 28
18. Sample of Google Trends Terms and their Principal Component (PC) Loadings 29

TABLES

1. Publication and Nowcast Lags by Indicator 7
2. Summary of Satellite-based Indicators: Frequency, Timeliness, and Sources 14
3. Training and Test Set Coverage by Indicator 23
4. Agriculture GVA Nowcasting Model Performance 25
5. Tourism Revenues Nowcasting Model Performance 26
6. Unemployment Rate Nowcasting Model Performance 28

Glossary

ARIMA	Auto Regressive Integrated Moving Average
ARIMAX	Auto Regressive Integrated Moving Average with exogenous regressors
AIS	Automatic Identification System
API	Application Programming Interface
ASI	Agricultural Stress Index
EN	Elastic Net
EVI	Enhanced Vegetation Index
GDP	Gross Domestic Product
GI	Google search-based Index
GT	Google Trends
GVA	Gross Value Added
MAE	Mean Absolute Error
MDI	Mean Decrease in Impurity
MIDAS	Mixed Data Sampling
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
OLS	Ordinary Least Squares
PLS	Partial Least Squares
RF	Random Forest
RMSE	Root Mean Squared Error
SIS	Sure Independence Screening
TCI	Temperature Condition Index
VCI	Vegetation Condition Index
VHI	Vegetation Health Index
VIP	Variable Importance in Projection
VS	Variable Selection
XGB	Extreme Gradient Boosting

Executive Summary

Access to reliable and timely data is critical for the quality of policy making. Non-traditional data sources can be used to generate accurate, real-time estimates of economic conditions, providing policymakers and analysts with timely data-driven insights ahead of official statistical releases.

This paper showcases how different types of non-traditional data—such as indices extracted from satellite imagery, Google Trends, and flight tracking information—can be leveraged to complement official statistics and monitor economic activity, and how these timely signals can be incorporated into nowcasting models to provide early estimates of key macroeconomic variables in Morocco. The approach is applied to (i) agricultural gross value added (ii) tourism revenues, and (iii) the unemployment rate.

The results demonstrate that the inclusion of non-traditional, high-frequency indicators into nowcasting models yields substantial improvements in forecast accuracy across all three indicators studied. Relative to benchmark models that rely solely on conventional time-series methods without access to non-traditional data, the enriched models improve the nowcasting results, as they reduce out-of-sample Root Mean Squared Error (RMSE) by 18 percent for agricultural gross value added (GVA), 14 percent for tourism revenues, and 25 percent for the unemployment rate, providing evidence of the informational value non-traditional data brings to real-time economic monitoring. In addition, machine learning models and statistical models are compared for each of the indicators, and it is found that whether machine learning or statistical models have higher predictive accuracy ultimately depends on the nature of the underlying series and the specific dynamics of each sector.

Beyond their application to Morocco, the methods and data sources demonstrated in this analysis are readily transferable to other countries. Given the near-global coverage of the non-traditional data sources used, the nowcasting frameworks developed in this paper can be replicated and adapted with minimal additional requirements. This scalability makes them particularly relevant across a wide range of countries to provide real-time insights before official statistics become available. More broadly, the results contribute to a growing evidence base on the utility of non-traditional data in macroeconomic forecasting and point to the wider potential of these approaches in strengthening economic monitoring in data-constrained environments.

Introduction

The strength of non-traditional data lies in its ability to capture economic signals before they appear in official statistics. Traditional macroeconomic statistics require time for collection, processing, and verification, therefore their publication is often with a lag and at lower frequency. This can constrain the ability of policymakers and analysts to assess economic conditions in real time, for instance at times of external shocks or during turning points in the business cycle. Non-traditional data sources—including satellite imagery, internet search queries, and flight tracking data, among others—serve as a valuable complement to official statistics, combining near real-time availability with a degree of granularity that conventional statistical systems cannot match. These alternative sources track signals that precede and accompany economic activity: whether it is vegetation conditions and temperatures measured by satellites, search patterns showing consumer interest, or ship movements indicating trade flows. Moreover, this data allows for a granular analysis of economic activity by region and sector. Incorporating non-traditional data into econometric models has been shown to improve nowcast accuracy across a range of indicators, particularly during rapid shifts when conventional data lags. These properties make non-traditional data a powerful tool for tracking economic conditions in real time, reducing the informational lags that traditionally constrain nowcasting and policy analysis.

While nowcasting has increasingly been applied to predict different macroeconomic variables, this paper contributes to the existing literature by using non-traditional data, which is available at higher frequency and shorter lags, rather than traditional data as inputs. Applying non-traditional data and nowcasting to the case of Morocco illustrates how this approach can be used. Morocco has high-quality official statistics that can be complemented with non-traditional data to shed light in real time on key sectors of the economy in the face of shocks.

- **Agriculture gross value added (GVA).**[1] Morocco's agricultural sector remains exposed to the risk of drought, with direct consequences for food prices—which account for 39 percent of the inflation basket—, rural employment, and overall economic stability. Given this volatility and the sector's significant macroeconomic influence, timely monitoring of agricultural conditions is of clear policy relevance. Leveraging alternative data sources, this paper estimates agricultural GVA growth starting around one month into the quarter, which is up to around five months before the official release (which is available 3 months after the end of the quarter), offering policymakers an early read on agricultural performance and its implications for inflation and employment.
- **Tourism revenues.** Tourism is a key growth driver and a sector of active strategic investment, with Morocco expanding air connectivity, upgrading infrastructure, and preparing to host major events including the 2025 Africa Cup of Nations (AFCON) and the 2030 FIFA World Cup. Monitoring the sector's performance in a timely manner is therefore of direct policy relevance. Using alternative data sources, this paper generates tourism revenue estimates 5 days after the month, which is around 3 weeks in advance of official figures (which are available 4 weeks after the end of the month), enabling a timelier assessment of how these investments are translated into economic activity.
- **Unemployment rate.** Unemployment remains a significant challenge for Morocco. Job losses in agriculture driven by drought conditions have weighed on employment, particularly in rural areas, while youth unemployment has continued to rise and has not been offset by sufficient job creation in

[1] Refers to the year-on-year growth rate of agricultural value added.

services and industry. In response, the authorities launched the Job Plan 2030, a roadmap targeting a reduction in the unemployment rate to 9 percent by 2030. Tracking progress against this objective in a timely manner is therefore of clear policy importance—and one where non-traditional data and nowcasting can add particular value, providing estimates of current unemployment conditions starting one month into the quarter, which is up to three months ahead of the official release (which is available one month after the end of the quarter).

This paper showcases the application of non-traditional data sources to complement official statistics and nowcast agricultural GVA, tourism revenues, and the unemployment rate. For each of the three indicators examined, relevant non-traditional data sources are identified and matched to the dynamics of the underlying series: vegetation and climate indices derived from satellite imagery for agriculture, travel-related search query data from Google Trends as well as flight tracking information from FlightRadar24 for tourism, and search query data capturing labor market dynamics from Google Trends for unemployment. The paper then shows how these indicators can be incorporated into nowcasting models to produce timelier estimates of agricultural GVA, tourism revenues, and the unemployment rate, with further details on the publication and nowcast lags presented in Table 1. Across indicators, models enriched with non-traditional data consistently outperform benchmarks that rely solely on conventional time-series methods—providing direct evidence of the informational value these sources bring to real-time economic monitoring. The paper further examines machine learning and statistical approaches, highlighting that model performance varies depending on the sector and the nature of the underlying data.

Table 1. Publication and Nowcast Lags by Indicator

Indicator	Publication Lag	Nowcast Lag
Agricultural GVA	~t+90 days (1 quarter after the end of the reference quarter)	~t-50 days (40 days into the reference quarter)
Tourism Revenues	~t+30 days (1 month after the end of the reference month)	~t+5 days (5 days after the end of the reference month)
Unemployment Rate	~t+30 days (1 month after the end of the reference quarter)	~t-60 days (30 days into the reference quarter)

Note: t denotes the end of the reference period.

The non-traditional data sources utilized in this paper are available with near-global coverage, updated at high frequency with minimum lags, and granular enough to allow disaggregated analysis at the sector or sub-national level. These sources are relevant across country income groups. For low-income and data-scarce economies, in cases where official data is unavailable, non-traditional data can serve as a proxy for economic activity[2]. For emerging markets and advanced economies, these sources can provide early estimates of different macroeconomic indicators and lead to significant time savings in generating estimates for policy makers, providing them with early insights into economic developments. The nowcasting framework demonstrated in this paper is also generalizable—the pre-processing pipeline, variable selection procedures, and model architecture can be adapted to different data environments and different structural characteristics. The framework accommodates different country contexts, target variables, and varying relationships between non-traditional indicators and the type of economic activity they track. From data extraction to nowcast

[2] See Sozzi (2024) for an application of satellite imagery and sensor-based indicators as proxies for economic activity in Somalia.

estimation, the end-to-end pipeline is designed as a replicable framework that can be tailored to different countries and sectors.

This paper contributes to the nowcasting literature by leveraging non-traditional data sources that are available at higher frequency and with shorter publication lags than conventional statistics, and by applying them to sector-specific outcomes. For agriculture, the contribution lies in focusing on agricultural gross value added—a sector that has received less attention in the nowcasting literature than aggregate output — and in combining satellite-based indicators such as vegetation and climate indices to capture the drivers of agricultural activity, which differ substantially from those of non-agricultural output. While satellite-based indicators are increasingly used in macroeconomic nowcasting, their application to sector-specific agricultural output remains limited. For tourism, the contribution lies in targeting tourism revenues—the indicator most directly relevant for external sector analysis and analyzing balance of payments dynamics—rather than the more commonly studied tourism-related indicators such as arrivals and overnight stays. Finally, for unemployment, while the use of Google Trends in nowcasting unemployment is well established in the literature—primarily in advanced economies—this paper applies the approach in an emerging market setting, demonstrating its relevance and applicability across different country contexts.

The paper is organized as follows: Section 2 briefly covers the relevant literature, Section 3 covers the data sources used and their relevance to monitoring the different indicators, Section 4 covers the methodology for nowcasting, Section 5 covers the results, and finally, Section 6 covers the conclusions and suggestions on improvement.

2. Literature Review

Non-traditional data is increasingly being used in monitoring economic activity and nowcasting models to produce timely estimates of different macroeconomic indicators. For example, Fotopoulou et al. (2026) address the near-complete absence of official GDP data in Venezuela by combining satellite-derived indicators, like nighttime lights, the Normalized Difference Vegetation Index (NDVI), the Enhanced Vegetation Index (EVI), and nitrogen dioxide emissions, with traditional macroeconomic data in a Random Forest nowcasting framework. Arslanalp et al. (2025) introduced a nowcasting model for global maritime trade that leverages satellite-based Automatic Identification System (AIS) data to track vessel movements in near real-time, providing a high-frequency proxy for global trade. Zheng et al. (2024) focused on nowcasting GDP growth rate and inflation expectation in China by integrating traditional macroeconomic data with novel textual data, showing that large-scale textual data capture rich qualitative signals that meaningfully enhance and complement traditional economic indicators. Austin et al. (2021) demonstrate how Google Places API and Google Trends data can be combined to develop high-frequency indicators of economic activity aligned with official statistical concepts and classifications. Woloszko (2020) introduced the OECD Weekly Tracker, a nowcasting tool designed to provide real-time estimates of weekly GDP growth by addressing the short historical depth of alternative data like Google Trends across 46 countries.

Recent literature has also explored the use of non-traditional data sources to nowcast agricultural gross value added, offering more timely and granular insights than conventional statistics. Bravo Higuera et al. (2024) developed a nowcasting estimator of Colombia's agricultural output by combining traditional statistics with big-data signals from Google Trends and Google News. Kaustubh et al. (2024) seek to strengthen sectoral GVA

nowcasting for India and find that incorporating a digital-activity index and a supply-side disruption index meaningfully improves post-pandemic nowcast performance.

The use of alternative data has also become central to real-time monitoring of tourism trends. Crispino and Mariani (2025) develop a nowcasting framework for tourist overnight stays in Italy combining payment card data and Google Search indices, showing that payment data is very useful in predicting tourism volumes. Bespalova (2022) develops an ARIMAX forecasting model for monthly tourism arrivals to Aruba, incorporating Google Trends data on destination searches, flight capacity utilization, and US consumer demand indicators. Havránek and Zeynalov (2021) examine the predictive power of Google Trends for monthly tourist arrivals and overnight stays in Prague, finding that search data up to two months ahead of arrivals provide significant forecast improvements. Cevik (2020) investigates whether travel-related internet search queries improve the accuracy of tourist arrivals forecasting for The Bahamas, finding that a Google Trends-augmented model outperforms a standard ARIMA benchmark by around 30 percent and improves on a multivariate model with income and price variables by more than 20 percent.

Growing literature also employs alternative data to track labor market conditions. Costa et al. (2024) developed a real-time nowcasting framework for the Portuguese monthly unemployment rate using daily Google Trends data and MIDAS regressions, accounting for the official two-month publication lag. The nowcasts prove timelier than provisional estimates from the national statistics office, providing accurate information 30 to 60 days in advance. Borup and Schütte (2022) develop a machine-learning forecasting framework for U.S. employment growth using a large panel of Google Trends search terms. They show that Google search activity is a strong out-of-sample predictor of future employment growth, and that combining targeted predictor selection with nonlinear models such as Random Forest substantially improves predictive performance. Caperna et al. (2022) address the challenge of constructing search-based unemployment indicators in a multi-country, multi-language setting by proposing a two-step data-driven procedure that uses the Google Topics framework to retrieve language-specific queries associated with unemployment across all EU27 countries, then applies Random Forest with Boruta feature selection to identify the queries with the strongest predictive signal in each country. D'Amuri and Marcucci (2017) evaluated the predictive power of a Google search-based index (GI) in tracking job-related queries for forecasting the US monthly unemployment rate. The authors demonstrate that autoregressive models augmented with the GI consistently outperform traditional leading indicators like initial claims, survey-based expectations, and economic policy uncertainty indices, particularly at medium to long forecast horizons.

While the literature demonstrates the growing use of non-traditional data in improving the timeliness of economic monitoring, many applications continue to focus on aggregate indicators. In agriculture, nowcasting agriculture output remains at a relatively early stage, with only a limited number of applications compared to aggregate GDP nowcasting. In the tourism literature, studies have largely focused on arrivals and overnight stays, with comparatively fewer studies focusing on revenue-based measures. For labor markets, the use of Google Trends is well established, particularly in advanced economies, with comparatively fewer applications in emerging market contexts. Taken together, this suggests scope for further extending the use of non-traditional data across different sectoral indicators and country settings.

3. Non-Traditional Data

This section discusses the different types of data used for monitoring and nowcasting the three target variables: (i) agriculture value added, (ii) tourism revenues, and (iii) unemployment rate.

3.1 Agriculture Value Added

Satellite imagery consists of digital representations of the Earth's surface captured by sensors mounted on orbiting satellites, which record reflected or emitted electromagnetic radiation across multiple spectral bands. By combining these spectral bands, researchers can derive a range of standardized indices that quantify specific surface conditions relevant to agricultural and environmental monitoring. A key advantage of satellite-derived data relative to conventional statistical sources is its spatial granularity: observations are available at the pixel level, with modern sensors such as Sentinel-2 providing resolutions as fine as 10 meters across its primary spectral bands. In addition, these indices are available at a high temporal frequency. MODIS vegetation index products are derived from daily atmosphere-corrected surface reflectance and composited into 16-day products (Huete et al., 2002), while the Sentinel-2 constellation achieves a revisit period of approximately five days (ESA, 2024). These two properties make satellite-derived indices particularly well suited as high-frequency indicators in nowcasting frameworks where timeliness and geographic detail are especially valuable.

Some of the most widely used indicators include vegetation indices, such as the Normalized Difference Vegetation Index (Rouse et al., 1974) and FAO's Agricultural Stress Index (FAO, 2017; Van Hoolst et al., 2016) that can be used to monitor crop growth and health throughout the agricultural cycle, precipitation measures derived from satellite sensors to proxy water availability and planting conditions in rain-fed agricultural systems (Pradhan et al., 2022), and nighttime light intensity which have been widely used as a proxy for broader economic activity (Henderson et al., 2012). Such indicators are available with shorter lags than official statistics, making them well suited for real-time monitoring of agricultural conditions.

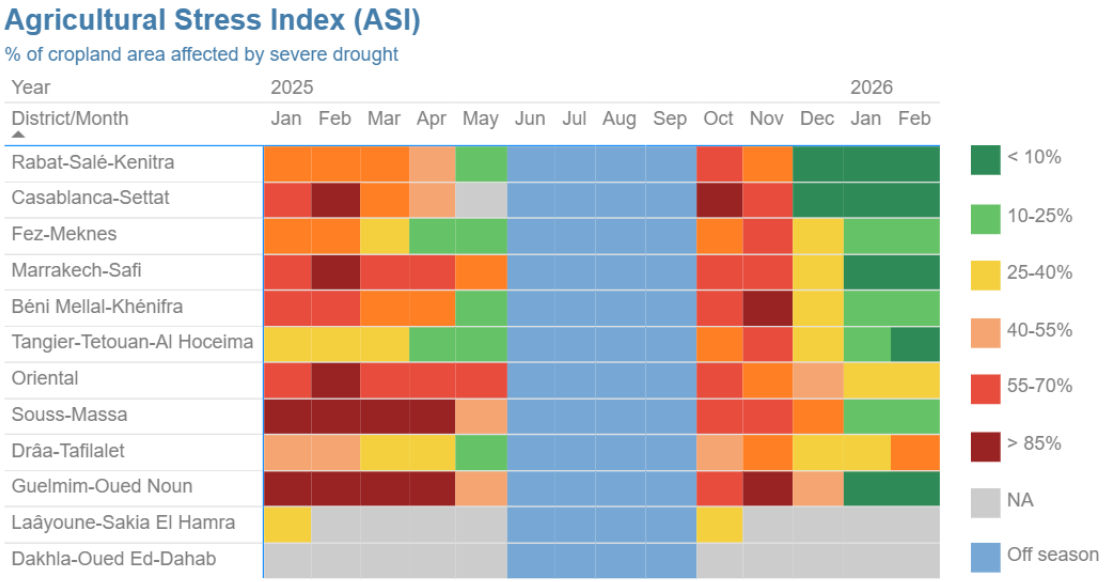
As these indicators capture crop conditions, vegetation health, and water availability, they provide timely signals of agricultural activity that can be informative for nowcasting agricultural value added. While the correlations between these indicators and agricultural output vary across variables, as shown in Figure 5, they indicate that these variables capture relevant aspects of agriculture dynamics, supporting their inclusion in the exploratory modeling phase. In this phase, a broad set of indicators is extracted to capture the full range of potentially relevant signals and subtle differences among related variables. The indicators explored are listed below, with further information on their frequency, timeliness, and sources in Table 2:

- **Agriculture Stress Index (ASI):** reports the share of cropland affected by water stress, offering a timely satellite-based indicator of cropland conditions that can serve as an early warning signal for drought conditions. By summarizing each month using the latest dekadal (10-day) ASI value, the Figure 1 heatmap highlights periods when stress peaks across districts. For instance, in early 2025, cropland stress appears elevated across regions, suggesting a broad deterioration in vegetation and moisture conditions—an early signal that drought impacts may be developing before they are fully reflected in official statistics. The indicator also captures the recovery in cropland conditions observed in early 2026, coinciding with increased rainfall following the end of the drought period. Used alongside

other climate indicators, ASI helps flag where and when conditions warrant closer monitoring and rapid response.

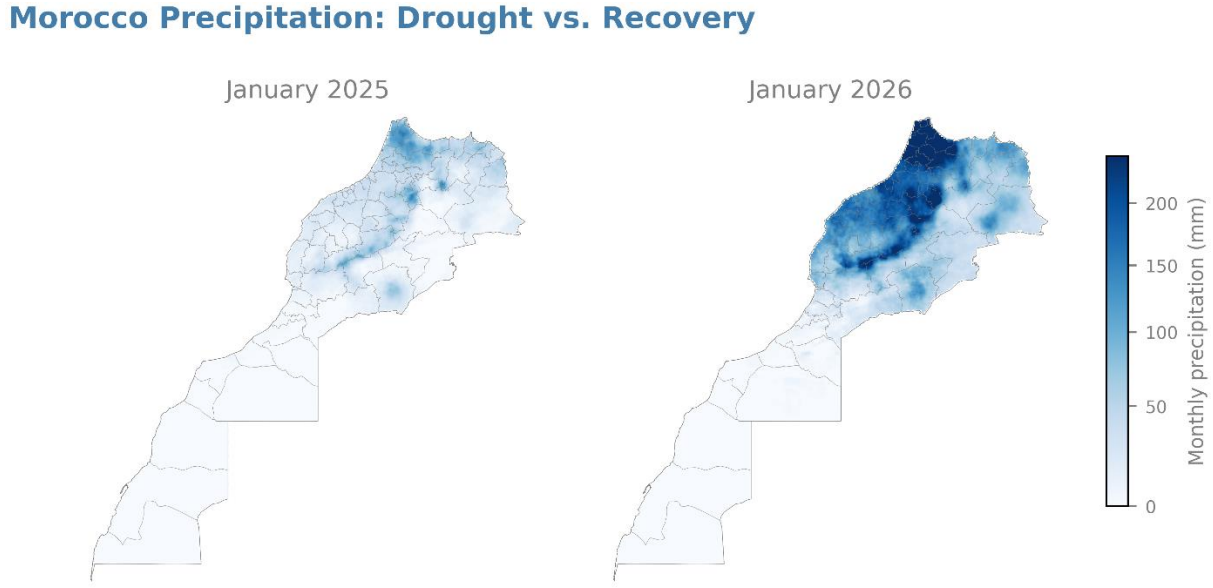
- **Precipitation:** an indicator of accumulated liquid and frozen water reaching the Earth's surface. It can serve as a proxy for water availability. Figure 2 highlights the large contrast between the drought conditions prevailing in January 2025 and the significant rainfall recovery recorded in January 2026. The spatial granularity of satellite data enables the identification of areas most severely affected by water deficits as well as those that experienced recovery following increased rainfall. This real-time tracking of precipitation dynamics is particularly valuable for agricultural nowcasting, as rainfall variability directly influences crop yields, vegetation health, and ultimately agricultural output.
- **Normalized Difference Vegetation Index (NDVI):** is an index that tracks the health of vegetation using red and near infra-red lights. It shows a high correlation with cereal production, as shown in Figure 3, suggesting that NDVI can provide a timely and granular proxy for agricultural output, and facilitate monitoring ahead of the annual official release.
- **Enhanced Vegetation Index (EVI):** a vegetation index designed to improve on NDVI by reducing atmospheric and background effects and improving sensitivity in areas with dense vegetation. It can be used as an indicator for vegetation conditions.
- **Vegetation Health Index (VHI):** a composite index that combines the Vegetation Condition Index (VCI) and the Temperature Condition Index (TCI) to assess vegetation health and monitor drought conditions.
- **Normalized Difference Vegetation Index (NDVI) Anomalies:** an indicator that measures deviations in NDVI from a historical average for the same period, helping identify unusually weak or strong vegetation conditions.
- **Temperature:** an indicator of air temperature at 2 meters above the surface. It can be used to assess the effect of heat on crop land.
- **Nighttime lights:** an indicator derived from low-light satellite sensors that measure nighttime radiance and are widely used as a proxy for economic activity. The nighttime lights shown in Figure 4 demonstrate how this data can be tracked at a regional level, showing higher average radiance in Morocco's main urban, industrial, and logistics centers. The comparison between 2015 and 2025 reveals a clear increase in luminosity, particularly in these areas, consistent with a growing concentration of urban and economic activity. While lighting upgrades can affect brightness, the pattern remains a useful, timely signal of regional activity.
- **Groundwater Storage Percentile:** an indicator showing the current groundwater storage relative to its historical distribution, expressed as a percentile. It can be used as an indicator for relative groundwater availability.

Figure 1. Agricultural Stress Index (ASI) Heatmap by District



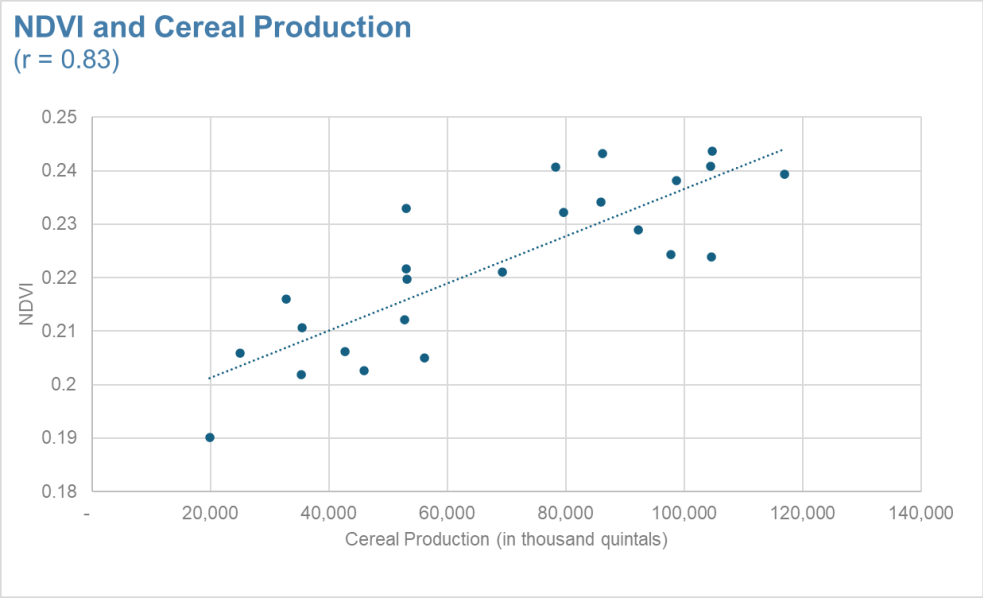
Sources: FAO and IMF staff Calculations
Note: FAO's Agricultural Stress Index (ASI) is reported at dekadal (10-day) frequency. For visualization only, we represent each month using the third dekad ASI value (days 21–end of month) as a proxy for monthly stress. All analysis and calculations use the original dekadal ASI data; the monthly series is provided solely for illustrative plotting and is not an official FAO monthly ASI product. The off-season refers to the period outside the active cropping cycle.

Figure 2. Precipitation in Morocco, January 2025 vs. January 2026



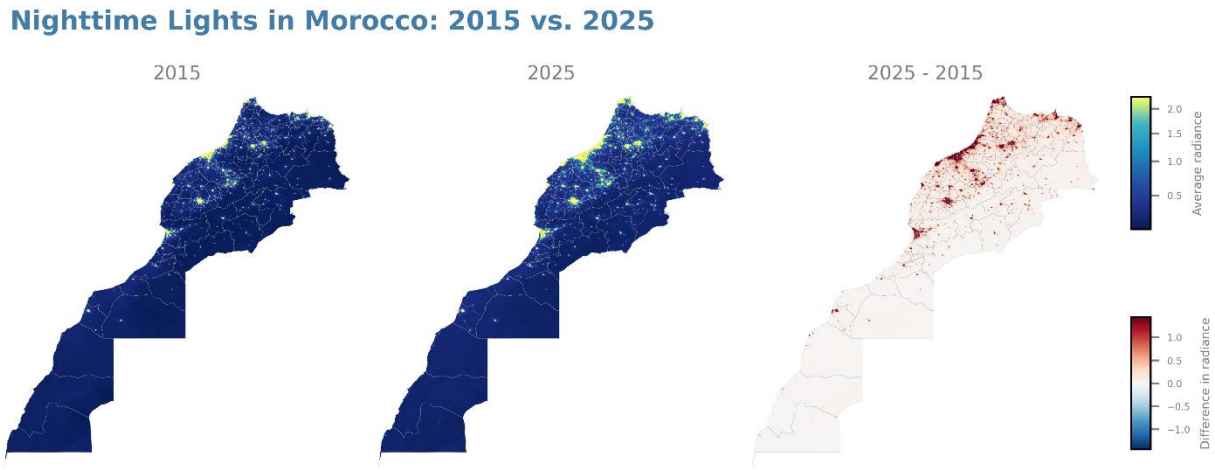
Sources: CHIRPS and IMF staff calculations.

Figure 3. Correlation between NDVI and Cereal Production



Sources: Ministry of Agriculture, NASA (MODIS), and IMF staff calculations.

Figure 4. Nighttime Lights, 2015 vs. 2025



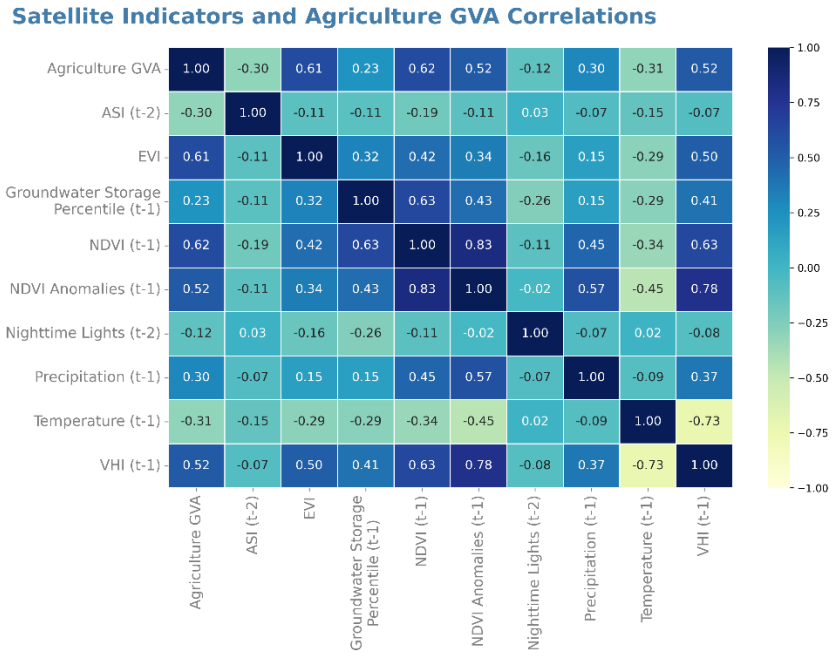
Sources: NASA and IMF staff calculations.
Note: Brighter colors indicate higher nighttime light intensity, which reflect areas with greater human and economic activity.

Table 2. Summary of Satellite-based Indicators: Frequency, Timeliness, and Sources

Indicator	Frequency	Timeliness	Source
Agriculture Stress Index (ASI)	Dekadal (10 days)	near real-time	Agricultural stress index near real time (Global - Dekadal - 1 km) - FAO ASIS
Precipitation	Daily	~5-6 days	ERA5-Land Daily Aggregated - ECMWF Climate Reanalysis
NDVI (Normalized Difference Vegetation Index)	8 days	~1-2 weeks	Landsat Collection 2 Tier 1 Level 2 8-Day NDVI Composite
Enhanced Vegetation Index	8 days	~1-2 weeks	Landsat Collection 2 Tier 1 Level 2 8-Day EVI Composite
Vegetation Health Index (VHI)	Dekadal (10 days)	near real-time	Vegetation health index near real time (Global - Dekadal - 1 km) - FAO ASIS
NDVI Anomalies	Dekadal (10 days)	near real-time	Non-Seasonal Indicators (Crops - Agricultural Stress Index System)
Temperature	Daily	~5-6 days	ERA5-Land Daily Aggregated - ECMWF Climate Reanalysis
Nighttime Lights	Daily	~1-2 weeks	VNP46A2: VIIRS Lunar Gap-Filled BRDF Nighttime Lights
Groundwater Storage Percentile	Weekly	~2-9 days	Groundwater storage percentile (GRACE-DA-DM Global v3.0, 0.25°, weekly)

Note: Timeliness figures are approximate and reflect typical availability under normal processing conditions. Actual data availability may vary due to satellite downtime, processing lags, reprocessing cycles, or differences between preliminary and final validated product releases.

Figure 5. Correlations between Satellite Indicators and Agriculture Value Added Growth Rates

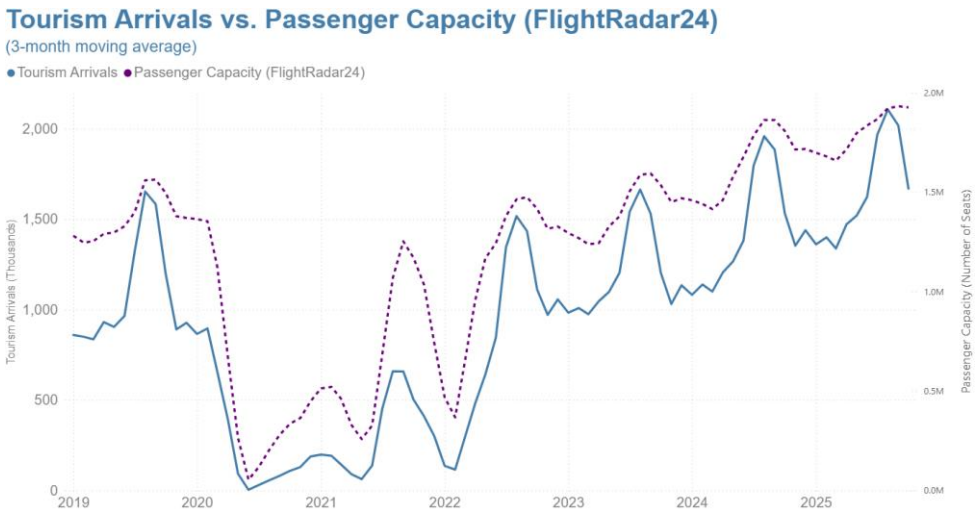


Sources: FAO, NASA, ECMWF, and IMF staff calculations.

3.2 Tourism Revenues

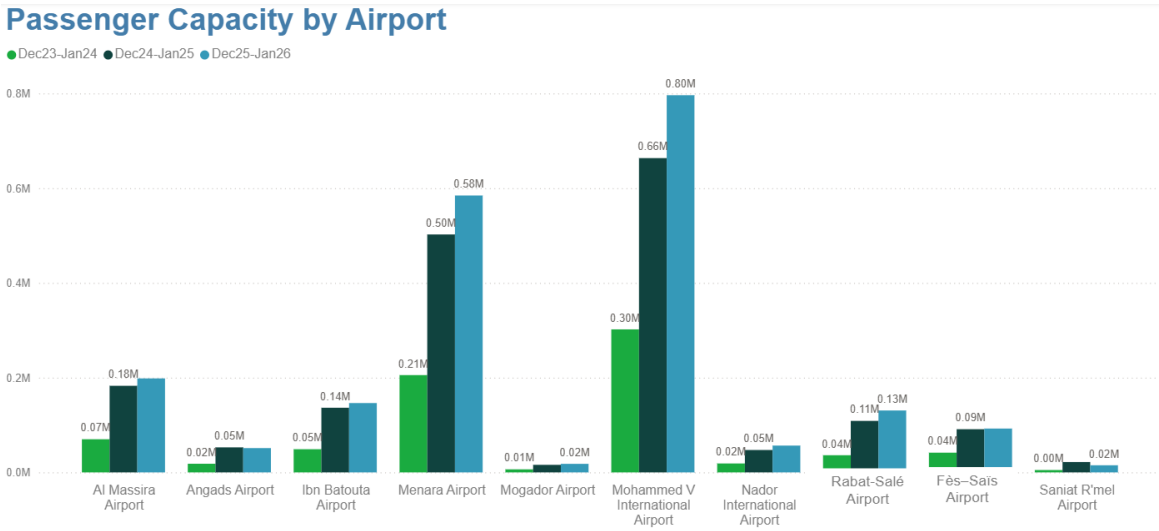
FlightRadar24 provides flight-tracking data such as aircraft movement and key flight details, such as origin, destination, and equipment type. The IMF Statistics Department developed the Passenger Capacity indicator derived from Flightradar24, a daily high-frequency indicator, available with a 5-day lag, designed to estimate the total seating capacity on flights by estimating the seat count from the aircraft type. The passenger capacity indicator shows strong co-movement with official tourism arrivals, with a correlation of 0.9 in levels and 0.95 when smoothed using a 3-month moving average, as shown in Figure 6 below. In addition, the data is disaggregated by country and airport, enabling the detailed tracking of increased tourism activity. Figure 7 demonstrates the increased tourism activity across major airports between December 2025 and January 2026, coinciding with the 2025 Africa Cup of Nations. These features make the passenger capacity a granular and timely proxy for tourism arrivals—available daily with a 5-day lag, while official arrivals data are released with a 1-month lag, and can in turn help inform the nowcast of tourism revenues, which are published with a 1-month lag. This is particularly valuable during peak events such as the AFCON 2025 and the 2030 FIFA World Cup.

Figure 6. Tourism Arrivals and Passenger Capacity (FlightRadar24)



Sources: Ministry of Tourism, FlightRadar24, and IMF staff calculations.

Figure 7. Breakdown of Passenger Capacity across Major Airports during the 2025 Africa Cup of Nations



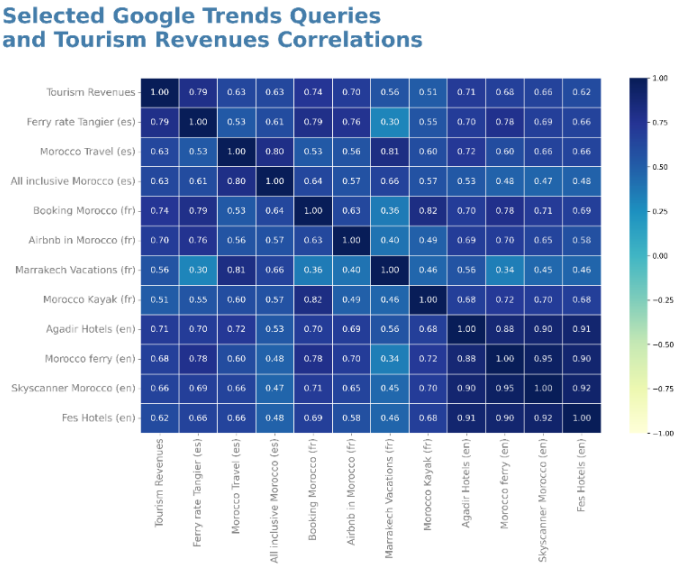
Sources: FlightRadar24 and IMF staff calculations.

Google Trends is a publicly available platform that reports the relative search intensity of user-defined query terms in Google Search, scaled from zero to one hundred[3] and available at daily, weekly, or monthly frequencies, depending on the query period and geographic scope. Choi and Varian (2012) showed that such search data can be useful for "predicting the present" as online search behavior provides timely information on current intentions, concerns, and economic activity before official statistics become available In the context of tourism, destination-related searches have been shown to be informative predictors of tourist arrivals (Havránek and Zeynalov, 2021). More broadly, Google Trends has been used in real-time economic monitoring, including in the OECD Weekly Tracker, which combines search data with machine learning techniques to produce timely estimates of GDP growth across 46 economies (Woloszko, 2020). A key advantage of Google Trends for nowcasting is its timeliness and high frequency relative to conventional indicators, with near real-time availability at a daily frequency, making it particularly suited for nowcasting applications.

Tourism revenues show strong notable associations with multiple Google Trends indices in different languages. Figures 8 and 9 illustrate this relationship through a correlation heatmap and selected time-series comparisons, highlighting the strong co-movement between search intensity and tourism activity. Combined with the established relevance of Google Trends for tracking tourism demand in the literature, these findings support the use of such indicators as timely signals for nowcasting tourism revenues.

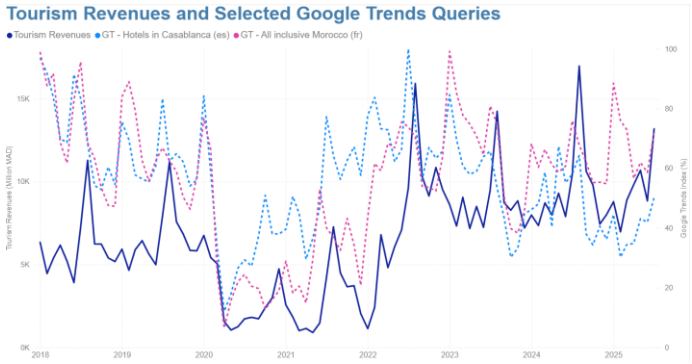
[3] Google Trends data are normalized and scaled from zero to one hundred, where one hundred represents the highest observed search interest for a given term within the selected time period and geographic region. All other values are expressed relative to this peak. The index represents relative search interest rather than absolute search volumes.

Figure 8. Tourism Revenues and Selected Google Trends Queries: Correlations



Sources: Ministry of Tourism, Google Trends, and IMF staff calculations.
Note: language abbreviations are as follows: en = English, es = Spanish, fr = French

Figure 9. Tourism Revenues and Selected Google Trends Queries: Time-Series Comparison



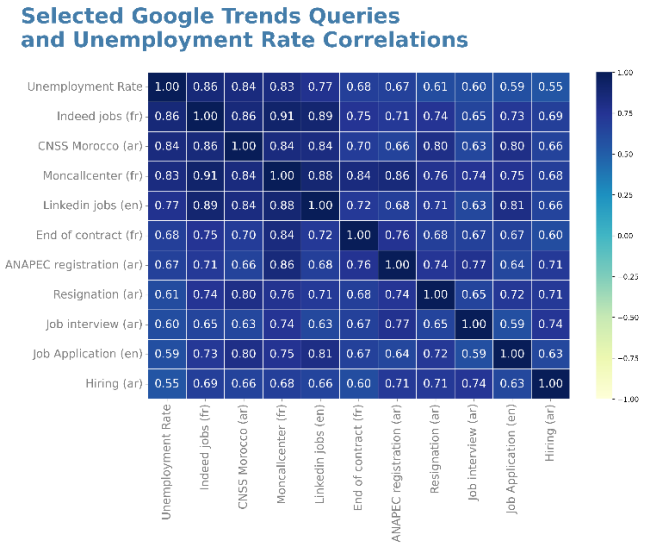
Sources: Ministry of Tourism, Google Trends, and IMF staff calculations.

3.3. Unemployment Rate

Google Trends also proves to be a useful data source for monitoring unemployment, as searches related to job hunting, employment platforms, and unemployment support can offer timely signals of labor-market conditions. D'Amuri and Marcucci (2017) demonstrate that Google job-search intensity improves the forecast of U.S. unemployment rate, while Askitas and Zimmermann (2009) show that internet search queries can be used to track and predict unemployment dynamics in Germany.

The unemployment rate is closely associated with several Google Trends indices related to job search activity. Figures 10 and 11 present this relationship through a correlation heatmap and selected time-series comparisons, revealing a notable co-movement between search activity and labor market conditions. Consistent with the established evidence in the literature of the usefulness of Google Trends in tracking unemployment dynamics, these findings reinforce the potential use of such indicators as timely inputs for nowcasting the unemployment rate ahead of its official release.

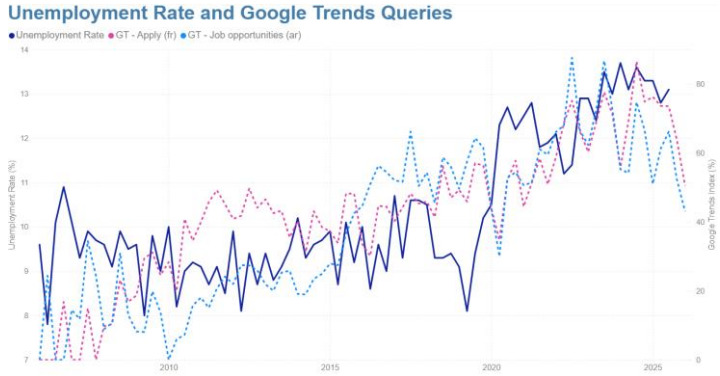
Figure 10. Unemployment Rate and Selected Google Trends Queries: Correlations



Sources: Google Trends and IMF staff calculations.

Note: language abbreviations are as follows: en = English, ar = Arabic, fr = French

Figure 11. Unemployment Rate and Selected Google Trends Queries: Time-Series Comparison



Sources: Google Trends and IMF staff calculations.

4. Methodology

This section outlines the methodology for the nowcasting framework, beginning with the data processing and transformation steps, followed by the specification of the modeling framework including the model selection, tuning, evaluation, and robustness checks.

4.1. Data Transformation and Processing

4.1.1. Agriculture Value Added

The target variable is the year-on-year (YoY) growth rate of quarterly agricultural GVA in volumes computed from official national accounts data. YoY growth rates are used to account for seasonality and stationarity and ensure consistency between the dependent and independent variables. Predictor variables are drawn from a suite of satellite-derived vegetation and climate-related indicators available at daily, weekly, or dekadal (10-day) frequencies. These high frequency indicators are aggregated to quarterly frequency, using sum for flow variables and mean for indices, and expressed in YoY growth rates to align with the target variable. For each satellite indicator, three temporal vintages are constructed: the contemporaneous quarter t, and two lags (t-1, t-2) to capture information lags and carry-over effects in vegetation and soil conditions. Prior to model estimation, highly correlated variables in the training set are filtered at a 90 percent correlation threshold to reduce redundancy and predictors are screened for data quality and stability.

4.1.2. Tourism Revenues

The target variable is monthly tourism receipts obtained from official statistics. Predictor variables comprise two complementary data sources: monthly Google Trends indices and the daily Passenger Capacity indicator derived from FlightRadar24 data, both aggregated to monthly frequency, using mean for Google Trends indices and sum for Passenger Capacity, to align with the target variable. Google Trends data were extracted once on April 3, 2026, using a fixed sample window and identical settings to ensure comparability of normalized (0–100) indices. To account for seasonality and stationarity, a log-difference transformation is applied to the Passenger Capacity, while a log(1+p) transformation is applied to Google Trends data, where p denotes the search interest index, to accommodate zero-valued observations in Google Trends data. A broad set of 100 tourism-related search terms are extracted, in French, Spanish, and English, reflecting key tourism origin markets, and covering a wide range of topics such as accommodation platforms, flights, and travel planning. This initial set is broad, following the approach of papers that work with a large candidate pool of Google Trends queries, before applying dimensionality reduction (Mulero and García-Hiernaux, 2021) since the subsequent pre-filtering pipeline and variable selection approach determines the number of variables that enter the model. Prior to model estimation, indices with excessive zeroes and highly correlated variables in the training set are filtered at a 90 percent correlation threshold to reduce redundancy. As Google Trends data can be sensitive to the choice of individual queries, this approach mitigates dependence on any specific term by relying on a broad candidate set combined with filtering and dimensionality reduction. For Google Trends indices, lag structures are evaluated: from contemporaneous (t) to six months (t-6).

4.1.3. Unemployment Rate

The target variable is the quarterly unemployment rate obtained from official statistics. Predictor variables consist of monthly Google Trends search indices capturing unemployment-related search activity. Google

Trends data were extracted once on April 3, 2026, using a fixed sample window and identical settings to ensure comparability of normalized (0–100) indices. Following the same approach outlined in the previous section, a broad set of 100 search terms related to unemployment and job-seeking behavior are extracted, in French, Arabic, and English. These terms cover multiple dimensions of unemployment, including vacancy listings, job-search platforms, unemployment benefits, and job application procedures. The multilingual approach is intended to capture variations in how search queries are conducted across the country. As Google Trends data can be sensitive to the choice of individual queries, this approach mitigates dependence on any specific term by relying on a broad candidate set combined with filtering and dimensionality reduction. Prior to model estimation, indices with excessive zeroes and highly correlated variables in the training set are filtered at a 90 percent correlation threshold to reduce redundancy. For Google Trends indices, the data is aggregated to quarterly frequency, using mean aggregation, to align with the target variable, and alternative lag structures from contemporaneous (t) to two quarters (t-2) are evaluated. To account for the structural break that occurred since Q4 2019, a dummy variable is included to capture the change in labor market dynamics in the context of the COVID pandemic.

4.2. Nowcasting Framework

This paper estimates and compares 12 model specifications across three families: linear models, tree-based ensemble methods, and time series models. The motivation is that the different indicators and data explored in this paper involve different types, frequencies, and predictor-target relationships. Linear models perform well when the relationship between the predictors and target is approximately linear and stable but may miss non-linearities. Tree-based methods can capture complex interactions without requiring an explicit functional form but are more data-intensive and may overfit on small samples. Time series models explicitly account for autoregressive dynamics in the target variable but require careful specifications of the exogenous components. To address the high dimensionality of the predictor set, each model is paired with a variable selection method to ensure that only the most informative predictors are retained, thereby reducing noise, and improving predictive performance. (Cashin et al., 2025; Jardet and Meunier, 2022). Finally, following standard practice in the nowcasting literature, a univariate ARIMA model estimated on the target variable alone is used as a benchmark to assess the incremental contribution of the non-traditional data sources to predictive accuracy.

4.2.1 Models

The model specifications explored in this study are as follows:[4]

- **Random Forest**
  An ensemble of independently grown decision trees, each trained on a bootstrap resample of the training data and a random subset of features at each split. Predictions are obtained by averaging across trees, which substantially reduces variance relative to a single tree. Hyperparameters include the number of trees, maximum tree depth, minimum samples per split and leaf node, and the size of the feature subset considered at each split.

- **Random Forest with Variable Selection**
  In this study, variable selection is performed using the Mean Decrease in Impurity (MDI) importance scores produced by a preliminary random forest fitted on the training data. MDI measures each

[4] See Breiman (2001) for Random Forest, Chen and Guestrin (2016) for XGBoost, Zou and Hastie (2005) for Elastic Net, Hastie et al. (2009) for Partial least Squares, Tibshirani (1996) for the LASSO variable selection applied to Ordinary Least Squares, and Box and Jenkins (1976) for ARIMA and ARIMAX.

feature's average contribution to reducing node impurity across all trees and splits. Features whose importance exceeds the mean MDI across all candidates are retained, and the tuned RF procedure is then applied on this reduced feature set. The selection step is performed exclusively on training data to avoid lookahead bias.

- **XGBoost**
  A gradient-boosted tree ensemble in which trees are added sequentially to correct the residuals of the current ensemble, with each tree regularized by constraining its depth and applying a shrinkage factor (learning rate) to its contribution. Unlike random forests, boosting is an iterative procedure in which trees are not independent, giving the method capacity to model complex interactions at the cost of a greater tendency to overfit with excessive iterations.

- **XGBoost with Variable Selection**
  In this study, variable selection is performed using the gain-based feature importance native to XGBoost. Gain measures the average improvement in the loss function brought by a feature across all splits in which it is used, making it a direct measure of predictive contribution rather than frequency of use. Features whose gain exceeds the mean gain across all candidates are retained, and the final model is estimated on this reduced feature set.

- **Elastic Net**
  A penalized linear regression that combines an L1 (LASSO) and L2 (Ridge) penalty on the coefficient vector. The mixed penalty simultaneously performs variable selection and coefficient shrinkage, making it well-suited to settings where many predictors are correlated. Two hyperparameters govern the penalty: the overall regularization strength α and the mixing parameter $\rho \in [0, 1]$ that interpolates between pure Ridge ($\rho = 0$) and pure LASSO ($\rho = 1$).

- **Elastic Net with Variable Selection**
  In this study, variable selection is performed by first estimating an Elastic Net model on the full predictor set and retaining only predictors with non-zero coefficients. The model is then re-estimated on the reduced set, yielding a post-selection specification that is more parsimonious.

- **Partial Least Squares**
  A dimension-reduction regression that constructs a small number of latent components as linear combinations of the predictors chosen to maximize covariance with the target. Unlike principal components, PLS components are supervised and therefore directly relevant to the outcome. The number of retained components is the main tuning parameter.

- **Partial Least Squares with Variable Selection**
  In this study, variable selection is performed using Variable Importance in Projection (VIP) scores, which summarize each predictor's contribution to explaining the target variable across all retained latent components. A VIP score above 1 is the conventional threshold for relevance, as it indicates that a variable contributes more than average to the model's explanatory power. Features with VIP greater than 1 are retained, and the model is refit on this reduced feature set.

- **Ordinary Least Squares**
  Ordinary Least Squares (OLS) is a standard linear regression that models the target variable as a linear function of the predictor variables, with parameters estimated by minimizing the sum of squared residuals.

- **Ordinary Least Squares with Variable Selection**
  In this study, the variable selection method performed is LASSO (Least Absolute Shrinkage and Selection Operator) regularization. The L1 penalty shrinks coefficients of less informative predictors to exactly zero. The penalty parameter λ is selected via k-fold cross-validation on the training set. Features corresponding to non-zero LASSO coefficients constitute the selected predictor set, upon which a standard OLS model is estimated.
- **Auto Regressive Integrated Moving Average with exogenous regressors (ARIMAX)**
  ARIMAX is an autoregressive integrated moving average model augmented with exogenous regressors, extending the standard ARIMA framework by incorporating external predictor variables alongside the autoregressive and moving average components of the target variable. The model is characterized by three key parameters: p (autoregressive order), d (degree of differencing), and q (moving average order). These parameters govern how past values of the target variable, differencing operations to achieve stationarity, and past forecast errors contribute to current predictions.
- **ARIMAX with Variable Selection**
  In this study, variables are ranked by their absolute marginal correlation with the target variable on the training set, following the Sure Independence Screening (SIS) approach of Fan and Lv (2008). This method is particularly suitable in high-dimensional settings, where including all predictors may introduce noise and lead to overfitting. The retained variables are then standardized and reduced via Principal Component Analysis (PCA) fitted on the training set to avoid lookahead bias, with the number of components chosen to explain 90 percent of the variance in the selected predictors. The resulting components are then used as regressors in the ARIMAX specification.
- **Auto Regressive Integrated Moving Average (ARIMA)**
  A pure time-series ARIMA model estimated on the target variable alone. This model serves as the primary benchmark against which the information value of non-traditional data is assessed. The lag orders are selected by exhaustive grid search over p, d, and q on the training set, with the specification minimizing the Akaike Information Criterion (AIC) retained as the benchmark. Models are estimated via maximum likelihood in a state-space representation, with non-convergent specifications discarded.

4.2.2. Hyperparameter Tuning and Model Evaluation

The dataset is split into a training set comprising 80 percent of the data, and a test set comprising the remaining 20 percent (Table 3). For models with tunable hyperparameters such as Random Forest, XGBoost, Elastic Net, and Partial Least Squares, the hyperparameters are selected within the training set using five-fold time-series cross-validation. The one-standard-error rule is applied to retain the simplest configuration whose root mean squared error (RMSE) falls within one standard error of the minimum, thereby reducing the risk of overfitting. (Breiman et al., 1984). The ARIMA baseline and ARIMAX model orders are selected by minimizing the Akaike Information Criterion (AIC), subject to convergence checks.

Model performance is evaluated using RMSE which penalizes large forecast errors and is a dominant metric in the nowcasting literature, as well as relative performance to the benchmark ARIMA model. To assess genuine out-of-sample predictive performance, all models are evaluated on a rolling pseudo-out-of-sample basis, where at each test period, the model is re-estimated using all available data up to that point and then used to generate the nowcast, with the estimation window expanding over time. This approach avoids reliance on a single fixed evaluation point, and instead produces multiple forecast errors, providing a more reliable assessment of predictive performance across periods and better reflecting how nowcasting operates in practice as new data

arrives. (Tashman, 2000). Mean absolute error (MAE) is also reported as a complementary metric to assess the robustness of the results.

Table 3. Training and Test Set Coverage by Indicator

Indicator	Training Dataset (80 percent)	Testing Dataset (20 percent)	Nowcast
Agricultural GVA	2015Q1 - 2023Q3	2023Q4 - 2025Q4	2026Q1
Tourism Revenues	2019M11 - 2024M10	2024M11 - 2026M2	2026M3
Unemployment Rate	2006Q1 - 2021Q4	2022Q1 - 2025Q4	2026Q1

4.2.3. Robustness Checks

To assess the stability of the results and robustness of model performance, three checks are conducted. First, the test period is divided into two equal sub-samples and model performance is evaluated separately in each half. Consistency in the ranking of the best-performing model across sub-periods provides evidence that the results are not driven by a particular phase of the economic cycle or a specific set of observations. Second, the nowcasting framework has been updated iteratively over the course of this study as new observations become available, with the best-performing model in each sector retaining its ranking across updates, providing real-time confirmation that the models reflect stable predictive performance. Finally, to assess whether the best-performing model provides a statistically significant improvement over the benchmark, the Clark–West test is conducted on pseudo out-of-sample forecasts.

5. Results

All models described in section 4.2.1 are estimated for each variable and compared based on rolling pseudo-out-of-sample RMSE. For agricultural GVA nowcasting using satellite-based indices, Random Forest delivers the lowest out-of-sample error. Satellite-based indicators such as vegetation indices and nighttime lights exhibit complex nonlinear relationships that machine learning models are well-positioned to capture. This finding is consistent with IMF work showing that incorporating satellite data in a Random Forest model, improves nowcasting accuracy, including applications to Cambodia (Maduako et al., 2026) and Venezuela (Fotopoulou et al., 2026), and that machine learning, more broadly, when integrating high frequency non-traditional indicators, delivers accurate and timely GDP estimates (Polo et al., 2025).

For tourism revenues nowcasting using Google Trends indices and the FlightRadar24 passenger capacity, the best performing model is ARIMAX with PCA-based dimensionality reduction of the Google Trends terms and the passenger capacity as exogenous regressors. This is consistent with evidence that incorporating web-based searches into a SARIMAX-type framework can outperform selected univariate and machine learning models for nowcasting tourism demand. (Lee, 2025).

For unemployment rate nowcasting using Google Trends indices, the best performing model is also ARIMAX with PCA-based dimensionality reduction of the Google Trends terms as exogenous regressors.

For both tourism and unemployment, the ARIMAX specification with variable selection outperforms the specification that includes the full set of Google Trends predictors. This is in line with the nowcasting literature, which highlights that Google Trends data are often high-dimensional relative to the sample size, and that including the full set of predictors can introduce noise without improving nowcast accuracy. As emphasized by

Ferrara and Simoni (2023), preselection or dimensionality reduction is often necessary to identify the most relevant predictors and obtain more stable and accurate nowcasts.

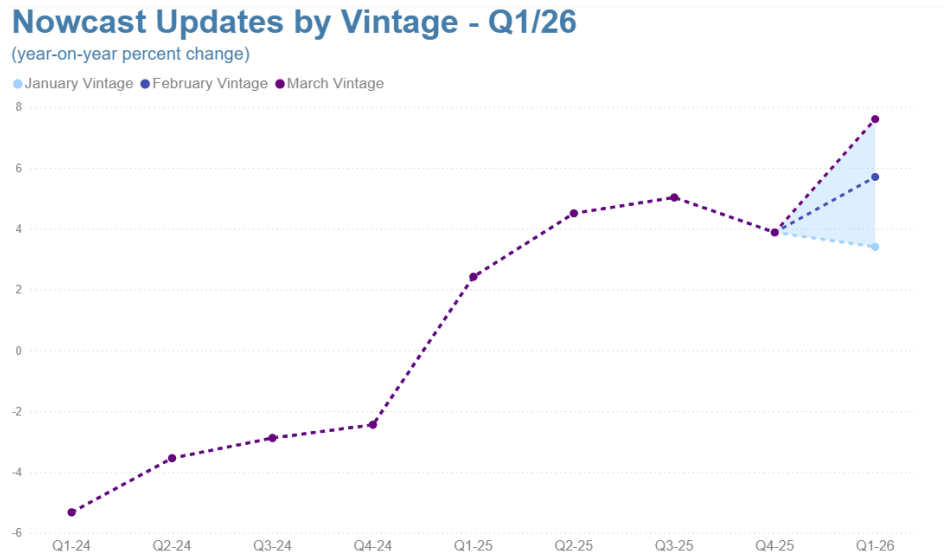
The Clark–West test indicates statistically significant improvements at the 5 percent level over the benchmark across all three indicators, providing supportive evidence of improved predictive performance. However, given the short evaluation windows, the strength of this evidence should be interpreted with caution.

These results suggest that the model selection should be guided by the nature of the underlying data. Machine learning models are well suited for leveraging satellite-based indicators with complex nonlinear structures, while statistical time-series models perform well when relationships are more stable and can be effectively captured through structured dynamics and exogenous inputs.

5.1. Agriculture Value Added

All models described in section 4.2.1 are estimated and compared based on rolling pseudo-out-of-sample RMSE. The final predictor set, following the initial pre-selection comprises EVI, NDVI anomalies, temperature, precipitation, and groundwater storage percentile, capturing key dimensions of vegetation health, climate, and water availability. Results for nowcasting agriculture GVA indicate that Random Forest outperforms other statistical and machine learning models (Table 4), delivering an 18 percent improvement in RMSE over the ARIMA benchmark. This improvement highlights the value of satellite-based indicators for enhancing nowcasting performance in the agricultural sector. In addition, the model has the lowest RMSE and MAE, indicating that it performs well both in terms of average errors and sensitivity to larger forecast deviations. The model captures key turning points, including the rebound in agriculture in early 2026. Figure 12 illustrates how the nowcast for 2026Q1 increases which each new data release as the quarter advances, capturing the improved agricultural conditions, such as higher precipitation and vegetation growth, ahead of official data releases expected in end-June 2026. More broadly, Figure 13 highlights the model's ability to track rapid shifts in the highly volatile agriculture sector.

Figure 12. Agriculture GVA Q1/26 Nowcasts by Vintage

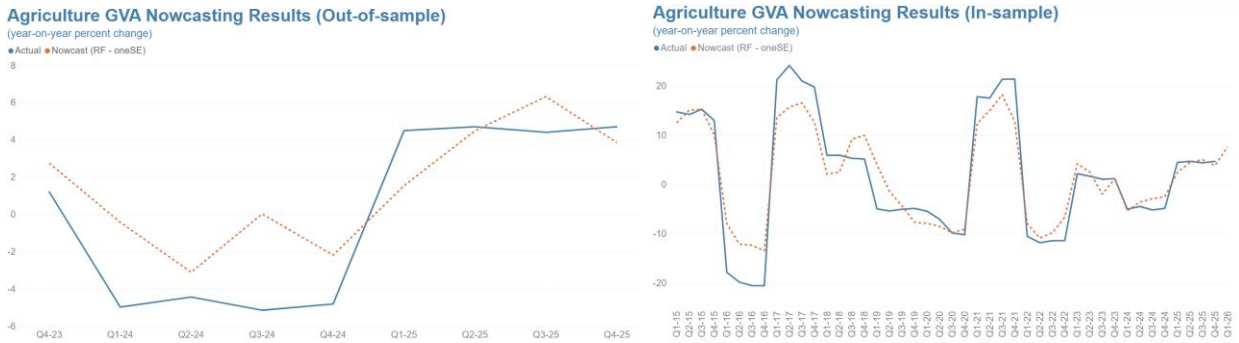


Sources: FAO, NASA, ECMWF, and IMF staff calculations.

Table 4. Agriculture GVA Nowcasting Model Performance

Model	Model Family	RMSE	MAE
RF	Tree-based	2.82	2.35
ARIMA	Time series	3.46	2.68
ARIMAX_VS	Time series	5.06	3.41
RF_VS	Tree-based	6.07	4.33
ARIMAX	Time series	6.20	5.46
XGB	Tree-based	6.94	6.44
XGB_VS	Tree-based	6.94	6.44
EN	Linear	8.03	5.18
PLS_VS	Linear	8.08	6.51
PLS	Linear	9.20	7.68
EN_VS	Linear	9.26	6.93
OLS	Linear	12.39	10.77
OLS_VS	Linear	12.39	10.77

Figure 13. Agriculture GVA Nowcasting Plots



Sources: Haut Commissariat au Plan (HCP), FAO, NASA, ECMWF, and IMF staff calculations.

Figure 14 shows the contribution of each indicator to the latest nowcast (2026Q1). The Shapley decomposition indicates that the nowcast is primarily driven by strong vegetation signals, across several months, followed by temperature and groundwater storage percentile. This highlights how the model combines different sources of information when constructing its estimate. It is important to note that the Shapley decomposition demonstrates the relative contribution of each feature to the prediction, but it does not imply causal relationships.

Figure 14. Shapley Decomposition of Q1/26 Nowcast



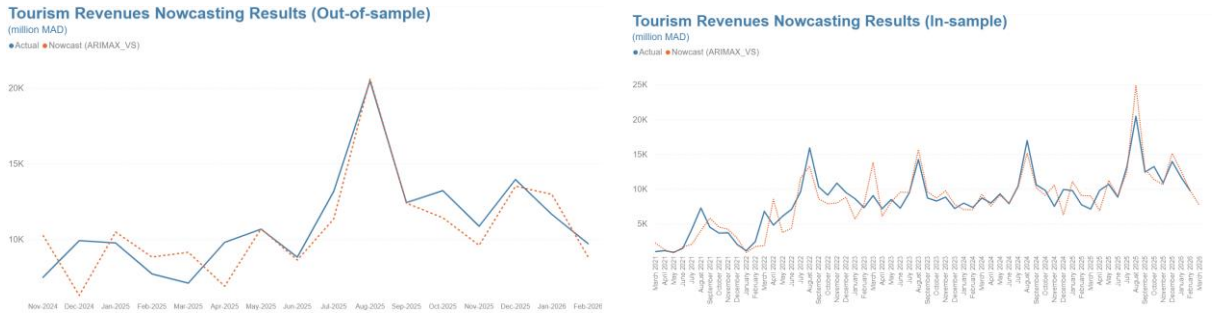
5.2. Tourism Revenues

All models described in section 4.2.1 are estimated and compared based on rolling pseudo-out-of-sample RMSE. For tourism revenues, contemporaneous Google Trends signals are found to outperform lagged specifications, suggesting that search activity within a given month captures information about tourism demand that was realized within the same reporting period, before official statistics become available. Results for nowcasting tourism revenues show that the best performing model is ARIMAX with Principal Component Analysis (PCA) based dimensionality reduction, retaining two principal components extracted from the pre-selected Google Trends series, alongside the passenger capacity calculated from FlightRadar24, as a direct exogenous regressor. The model achieves a 14 percent improvement in RMSE relative to the ARIMA baseline (Table 5), providing evidence of the informational value of non-traditional data for nowcasting tourism revenues. In addition, the model has the lowest RMSE and MAE, indicating its robustness and consistency across benchmarks. The model captures the overall impact of the 2025 Africa Cup of Nations (AFCON) event, with slight smoothing of the effect across December 2025 and January 2026, as demonstrated in Figure 15.

Table 5. Tourism Revenues Nowcasting Model Performance

Model	Model Family	RMSE	MAE
ARIMAX_VS	Time series	0.19	0.14
ARIMA	Time series	0.22	0.16
RF	Tree-based	0.22	0.17
RF_VS	Tree-based	0.23	0.18
XGB	Tree-based	0.28	0.21
XGB_VS	Tree-based	0.28	0.21
PLS	Linear	0.32	0.25
PLS_VS	Linear	0.43	0.35
EN_VS	Linear	0.46	0.38
EN	Linear	0.48	0.39
OLS_VS	Linear	0.55	0.44
OLS	Linear	2.53	1.44
ARIMAX	Time series	2.73	2.04

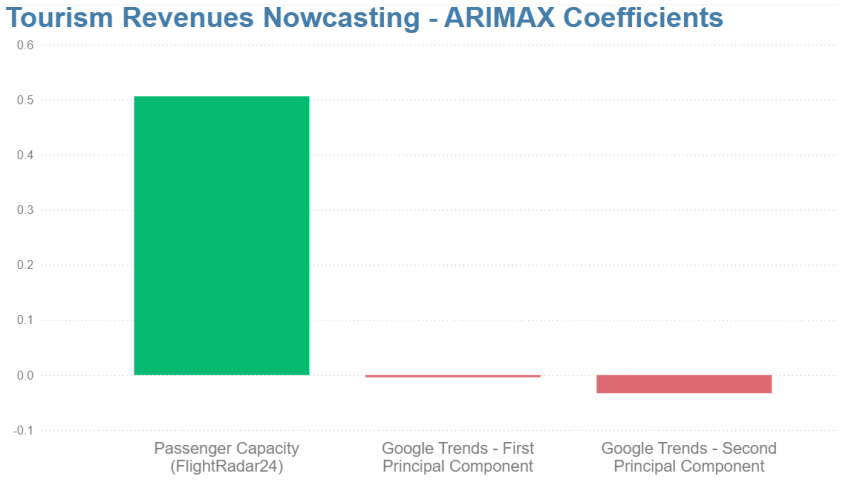
Figure 15. Tourism Revenues Nowcasting Plots



Sources: Ministry of Tourism, Google Trends, FlightRadar24, and IMF staff calculations.

Figure 16 presents the estimated coefficients of the exogenous regressors in the ARIMAX model for tourism revenues nowcasting. The passenger capacity, computed from FlightRadar24 data, has the largest coefficient (0.5), indicating that it is the dominant driver of tourism revenue dynamics in the model. The Google Trends principal components contribute more modestly, however, models estimated without Google Trends data yield higher out-of-sample errors, indicating that internet search data provides incremental predictive power beyond what flight capacity alone can capture, highlighting the informational value and complementarity of both sources in nowcasting tourism revenues.

Figure 16. Tourism Revenues Nowcasting ARIMAX Model Coefficients



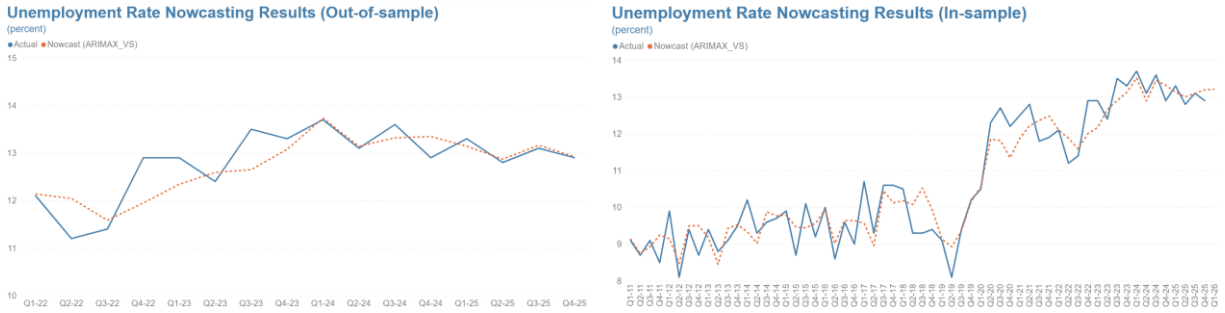
Sources: Google Trends, FlightRadar24, and IMF staff calculations.

5.3 Unemployment Rate

All models described in section 4.2.1 are estimated and compared based on rolling pseudo-out-of-sample RMSE. For unemployment rate, contemporaneous Google Trends signals also outperform the lagged specifications, suggesting that job-related behavior reflects current market conditions in real time. The best performing model is ARIMAX with Principal Component Analysis (PCA) based dimensionality reduction, retaining six principal components, extracted from the pre-selected Google Trends series, alongside a dummy variable capturing the structural break following COVID as exogenous regressors. The model achieves a

25 percent improvement in RMSE relative to the ARIMA baseline (Table 6), providing evidence of the informational value of Google Trends data for nowcasting unemployment rate. The best-performing model also achieves the lowest values on both RMSE and MAE, indicating robustness across metrics. The model is able to capture the downward trend in unemployment following the end of the drought period, as demonstrated in Figure 17.

Figure 17. Unemployment Rate Nowcasting Plots



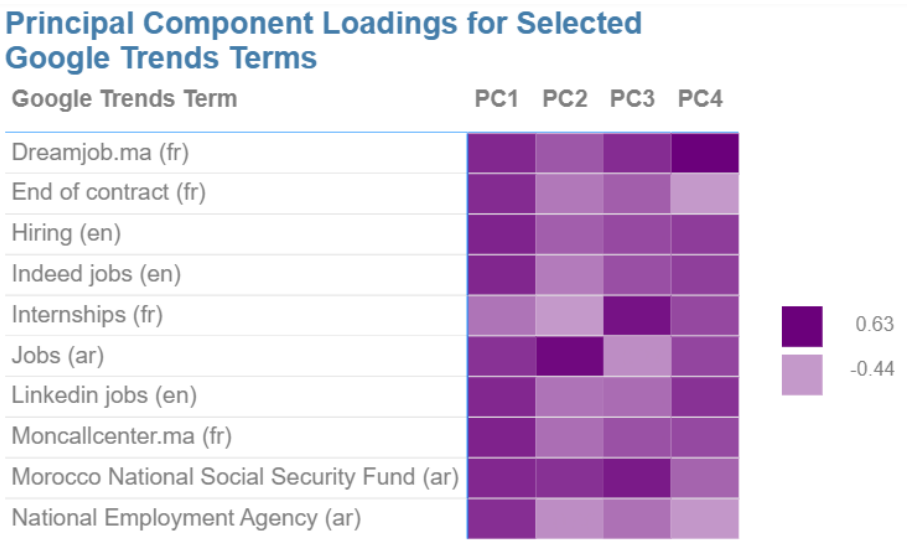
Sources: Google Trends, and IMF staff calculations.

Table 6. Unemployment Rate Nowcasting Model Performance

Model	Model Family	RMSE	MAE
ARIMAX_VS	Time series	0.44	0.31
ARIMA	Time series	0.59	0.45
RF_VS	Tree-based	0.69	0.55
RF	Tree-based	0.72	0.58
OLS_VS	Linear	0.79	0.68
PLS_VS	Linear	0.85	0.74
XGB	Tree-based	0.92	0.74
XGB_VS	Tree-based	0.92	0.74
PLS	Linear	1.05	0.95
EN	Linear	1.64	1.50
EN_VS	Linear	1.73	1.60
OLS	Linear	1.99	1.51
ARIMAX	Time series	2.20	1.86

Figure 18 presents a sample of the selected Google Trends terms and their loadings across the first four principal components. Each component combines terms across languages and captures distinct patterns related to labor market activity, such as job applications, resignations, internships, and unemployment. This highlights how diverse search behavior can be aggregated into meaningful signals that reflect underlying labor market dynamics. It is important to note that the loadings show how each term contributes to the principal component, but they are not directly interpretable as a causal or predictive of unemployment.

Figure 18. Sample of Google Trends Terms and their Principal Component (PC) Loadings



Sources: Google Trends and IMF staff calculations.
Note: Higher (more positive) principal component (PC) loadings indicate that the term contributes more strongly in the direction of that component, while lower (more negative) loadings indicate contribution in the opposite direction. Language abbreviations are as follows: fr = French, ar = Arabic, and en = English.

6. Conclusions

This paper demonstrates that non-traditional data sources can meaningfully complement official statistics to produce more timely and more accurate nowcasts of key macroeconomic indicators, applied to the case of Morocco. Non-traditional data is available at high-frequency and with short lags and can therefore complement official statistics. Across the three indicators analyzed, models that incorporate non-traditional data consistently surpass benchmarks based only on standard time-series techniques, demonstrating the added informational value these sources contribute to real-time economic monitoring. The performance advantage of machine learning compared to traditional statistical models varies by sector and type of data, with no single method outperforming the others across all indicators.

The data sources and nowcasting framework developed in this paper are explicitly designed for replicability in other countries. The framework can be adapted to the sectors most relevant in each country's context, with non-traditional data sources selected and matched to the specific dynamics of each sector. This flexibility means that while the application here focuses on agriculture, tourism, and unemployment in Morocco, the same approach could be applied to industry, construction, retail, or any other sector for which timely monitoring is a priority.

Further work could explore additional types of non-traditional data to monitor economic activity, broadening the scope of what can be tracked in real time. On the modelling side, future research could investigate additional machine learning and statistical approaches to assess whether alternative model configurations yield further accuracy gains. Finally, this pipeline can be leveraged across other countries where data is scarce, providing policymakers with more timely estimates of key economic indicators to support faster and better-informed policy decision-making.

References

Austin, P., Marco Marini, A. Sanchez, C. Simpson-Bell, and J. Tebrake. 2021. *Using the Google Places API and Google Trends Data to Develop High-Frequency Indicators of Economic Activity*. IMF Working Papers 2021/295. Washington, DC: International Monetary Fund.

Arslanalp, S., S. Mo Choi, P. Kamali, R. Koepke, M. McKetty, M. Ruta, M. Saraiva, A. Sozzi, and J. Verschuur. 2025. *Nowcasting Global Trade from Space*. IMF Working Papers 2025/093. Washington, DC: International Monetary Fund.

Askitas, N., and K. F. Zimmermann. 2009. *Google Econometrics and Unemployment Forecasting*. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1480251.

Bespalova, O. G. 2022. *Modeling and Forecasting Monthly Tourism Arrivals since the COVID-19 Pandemic: The Case of Aruba*. IMF Working Papers 2022/226. Washington, DC: International Monetary Fund.

Borup, D., and E. C. M. Schütte. 2022. "In Search of a Job: Forecasting Employment Growth Using Google Trends." *Journal of Business and Economic Statistics* 40 (1): 186–200.

Box, G. E. P., and G. M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. Revised edition. San Francisco: Holden-Day.

Bravo Higuera, D. F., L. D. Parra Bernal, M. L. Argote Cusi, and G. A. Torres Pineda. 2024. "Colombian Agricultural Sector's Early Estimator of Gross Domestic Production Using Nowcasting and Big Data Methods." *Journal of Technology Management and Innovation* 19 (2): 54–66. https://doi.org/10.4067/S0718-27242024000200054.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.

Caperna, G., M. Colagrossi, A. Geraci, and G. Mazzarella. 2022. "A Babel of Web Searches: Googling Unemployment during the Pandemic." *Labour Economics* 74: 102097.

Cashin, P., F. Han, I. Sabuga, J. Xie, and F. Zhang. 2025. *Parameter Proliferation in Nowcasting: Issues and Approaches—An Application to Nowcasting China's Real GDP*. IMF Working Papers 2025/217. Washington, DC: International Monetary Fund.

Cevik, S. 2020. *Where Should We Go? Internet Searches and Tourist Arrivals*. IMF Working Papers 2020/022. Washington, DC: International Monetary Fund.

Chen, T., and C. Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. New York: Association for Computing Machinery.

Choi, H., and H. Varian. 2012. "Predicting the Present with Google Trends." *Economic Record* 88 (s1): 2–9.

Costa, E. A., M. E. Silva, and A. B. Galvão. 2024. "Real-Time Nowcasting of Monthly Unemployment Rates with Daily Google Trends Data." *Socio-Economic Planning Sciences* 95: 101963. https://doi.org/10.1016/j.seps.2024.101963.

Crispino, M., and V. Mariani. 2025. "A Tool to Nowcast Tourist Overnight Stays with Payment Data and Complementary Indicators." *Italian Economic Journal* 11 (1): 285–312.

D'Amuri, F., and J. Marcucci. 2017. "The Predictive Power of Google Searches in Forecasting U.S. Unemployment." *International Journal of Forecasting* 33 (4): 801–816. https://doi.org/10.1016/j.ijforecast.2017.03.004.

European Space Agency (ESA). 2024. *Sentinel-2* https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2.

Fan, J., and J. Lv. 2008. "Sure Independence Screening for Ultrahigh Dimensional Feature Space." *Journal of the Royal Statistical Society: Series B* 70 (5): 849–911.

Ferrara, L., and A. Simoni. 2023. When Are Google Data Useful to Nowcast GDP? An Approach via Preselection and Shrinkage. Journal of Business & Economic Statistics 41 (4): 1188–1202.

Food and Agriculture Organization of the United Nations (FAO). n.d. *Agricultural Stress Index System (ASIS)*. Retrieved March 1, 2026, from http://www.fao.org/giews/earthobservation/.

Fotopoulou, E., I. Maduako, M. B. Sbrancia, and P. Srivastava. 2026. *Nowcasting Economic Growth with Machine Learning and Satellite Data*. IMF Working Papers 2026/020. Washington, DC: International Monetary Fund.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer.

Havránek, T., and A. Zeynalov. 2021. "Forecasting Tourist Arrivals: Google Trends Meets Mixed-Frequency Data." *Tourism Economics* 27 (1): 129–148.

Henderson, J. V., A. Storeygard, and D. N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira. 2002. "Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices." *Remote Sensing of Environment* 83 (1–2): 195–213.

Jardet, C., and B. Meunier. 2022. "Nowcasting World GDP Growth with High-Frequency Data." *Journal of Forecasting* 41 (6): 1181–1200.

Kaustubh, K., S. S. Bhadury, and S. Ghosh. 2024. "Reinvigorating GVA Nowcasting in the Post-Pandemic Period: A Case Study for India." *Bulletin of Monetary Economics and Banking* 27: 95–130.

Lee, G.-C. 2025. "A Data-Driven Approach to Tourism Demand Forecasting: Integrating Web Search Data into a SARIMAX Model." *Data* 10 (5): 73.

Maduako, I., D. Rijal, and A. Sanchez Rodelgo. 2026. *Satellite Data for Nowcasting: Estimating Cambodia's GDP in Real Time Using Machine Learning*. Selected Issues Paper 2026/001. Washington, DC: International Monetary Fund.

Mulero, R., and A. García-Hiernaux. 2021. "Forecasting Spanish Unemployment with Google Trends and Dimension Reduction Techniques." *SERIEs* 12 (3): 329–349.

Polo, G., Y. Gao Rollinson, Y. Korniyenko, and T. Yuan. 2025. *Nowcasting GCC GDP: A Machine Learning Solution for Enhanced Non-Oil GDP Prediction*. IMF Working Papers 2025/268. Washington, DC: International Monetary Fund.

Pradhan, R. K., Y. Markonis, M. R. Vargas Godoy, A. Villalba-Pradas, K. M. Andreadis, E. I. Nikolopoulos, S. M. Papalexiou, A. Rahim, F. J. Tapiador, and M. Hanel. 2022. "Review of GPM IMERG Performance: A Global Perspective." *Remote Sensing of Environment* 268: 112754.

Rouse, J. W., R. H. Haas, J. A. Schell, and D. W. Deering. 1974. “Monitoring Vegetation Systems in the Great Plains with ERTS.” In *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium*, Vol. I, 309–317. Washington, DC: National Aeronautics and Space Administration.

Sozzi, A. 2024. “Harnessing Satellite Data for Economic Monitoring in Fragile States: Application to Somalia.” In Somalia Selected Issues Paper. Washington, DC: International Monetary Fund.

Tashman, L. J. 2000. “Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review.” *International Journal of Forecasting* 16 (4): 437–450.

Tibshirani, R. 1996. “Regression Shrinkage and Selection via the Lasso.” *Journal of the Royal Statistical Society: Series B* 58 (1): 267–288.

Van Hoolst, R., H. Eerens, D. Haesen, A. Royer, L. Bydekerke, O. Rojas, Y. Li, and P. Racionzer. 2016. “FAO’s AVHRR-Based Agricultural Stress Index System (ASIS) for Global Drought Monitoring.” *International Journal of Remote Sensing* 37 (2): 418–439.

Woloszko, N. 2020. *Tracking Activity in Real Time with Google Trends*. OECD Economics Department Working Papers No. 1634. Paris: OECD Publishing.

Zheng, T., X. Fan, W. Jin, and K. Fang. 2024. “Words or Numbers? Macroeconomic Nowcasting with Textual and Macroeconomic Data.” *International Journal of Forecasting* 40 (2): 746–761.

Zou, H., and T. Hastie. 2005. “Regularization and Variable Selection via the Elastic Net.” *Journal of the Royal Statistical Society: Series B* 67 (2): 301–320.



PUBLICATIONS

Leveraging Non-traditional Data for Macroeconomic Nowcasting: The Case of Morocco

Working Paper No. WP/2026/108