

# Jitto Full Stack Engineering Challenge: A250

## Background

You're a data scientist at a global policy think tank. Your job is to monitor early warning signals of **societal unrest**—such as protests, strikes, or civil disturbances. You've been provided with a dataset that aggregates monthly indicators for 50 regions over a 10-year period.

Each month, for each region, you receive a set of economic, environmental, and sociopolitical indicators, along with whether any significant unrest occurred that month. Your task is to build a model that can **forecast the probability of unrest one month in advance**, based on these signals.

## Objectives

1. Build a predictive model that outputs the **probability** of unrest for each region-month.
2. Identify and explain which indicators are most influential in predicting unrest.
3. Evaluate your model's performance and calibration.
4. Propose simple, actionable alert rules a policy analyst could use.

## Dataset Overview

You will receive a file named: **unrest.csv**

Each row in the dataset represents a single region during a specific month.

## Columns

Column	Description
region_id	Unique identifier for each region
month	The first day of the month for the observation
gini_index	A measure of income inequality (0 = perfect equality, 1 = maximum inequality)
unemployment_rate	Percentage of the labor force that is unemployed
inflation_rate	Year-over-year change in consumer prices, in percentage points
media_sentiment_index	Sentiment score for regional news coverage (range: -1 = strongly negative, 1 = strongly positive)
rainfall_deviation	Deviation of current rainfall from historical average, expressed in standard deviations (z-score)
policy_instability_index	A rating from 0 to 3 indicating the level of political instability (0 = stable, 3 = high volatility)
prior_unrest_count	A decayed count of past unrest events in the same region over the previous few months
population_density	Number of people per square kilometer in the region
food_price_index	Relative cost of food goods (higher values suggest increased food costs or scarcity)
fuel_subsidy_cut	Indicator (0 or 1) for whether fuel subsidies were recently reduced or eliminated
gov_approval	Public approval of the government (range: 0 = no approval, 1 = full approval)
neighbor_unrest_prev	Number of nearby regions that experienced unrest in the previous month
unrest_event	Target variable: 1 if unrest occurred during the current month, 0 otherwise

## Constraints

Your solution must be written in **Python 3.12 or later**.

You may use the following libraries:

- **Core:** numpy, pandas, polars, scipy, matplotlib, networkx
- **Modeling:** scikit-learn, tensorflow, keras, pytorch
- **Other:** Any other open-source, Python-based library commonly used for data science is permitted—**except** tools that automate machine learning or forecasting pipelines. This includes but is not limited to: AutoGluon, H2O, TPOT, PyCaret, FLAML, Prophet, and similar frameworks.

Use of external APIs, pretrained models, or commercial platforms is not allowed. Your work should reflect your own modeling logic and analytical reasoning.

## Bonus (Optional)

Build a simple dashboard or web app that allows a user to:

- Select a region and view its predicted risk for the next month.
- Explore how indicators have evolved over time.
- Understand the key contributors to the model's output.

## Deliverables

All deliverables must be organized and published in a **public GitHub repository**.

Your repo must include:

- A Jupyter notebook with your full solution and analysis.
- Any scripts used to preprocess data, engineer features, or train models.
- A concise summary memo or write-up explaining your results and recommendations.
- (Optional) A working dashboard or interactive web app if you pursue the bonus section.

Make sure your README includes clear instructions to reproduce your results.

## What We're Looking For

Area	What Matters
Modeling	Accurate, well-calibrated predictions
Reasoning	Clear thinking about indicators and risk dynamics
Judgment	Thoughtful decisions, not just technical execution
Communication	Precise, concise, and actionable insights
Integrity	A realistic understanding of uncertainty and limitations

## Final Note

This challenge isn't about showing off libraries or tuning for leaderboard metrics. It's about your ability to reason under uncertainty, prioritize signal over noise, and translate complex patterns into clear insight.

Good luck !