Session 2: Hands-on - LSTM Drought Monitoring Lab

Mindanao Case Study with GEE-Sourced NDVI Forecasting

Stylianos Kotsopoulos

Session Agenda

- **Duration:** 2.5 hours (150 minutes)
- **Flow:**
  - 0-15: Kickoff & objectives
  - 15-45: Module 1 - Authenticate & fetch NDVI from GEE
  - 45-75: Module 2 - Tidy data & build sequences
  - 75-115: Module 3 - Configure PyTorch LSTM workflow
  - 115-135: Module 4 - Train & monitor
  - 135-145: Module 5 - Evaluate & drought decisions
  - 145-150: Reflection & wrap-up

Learning Objectives

- Retrieve Sentinel-2 NDVI time series for Mindanao provinces using Google Earth Engine
- Transform GEE exports into supervised sliding-window datasets (NDVI-only baseline)
- Train, tune, and monitor a PyTorch LSTM with strict temporal validation
- Translate forecasts into drought alarms via MAE/RMSE and NDVI thresholds

---

Resources

- Student notebook: `day4/notebooks/day4_session2_lstm_drought_lab_STUDENT.ipynb`
- Instructor notebook: `day4/notebooks/day4_session2_lstm_drought_lab_INSTRUCTOR.ipynb`
- Earth Engine Python API (`earthengine-api`) + Cloud project ID (set `EE_PROJECT`) + authentication (`ee.Authenticate(..., project=...)`)
- Optional fallback CSV (sample export): `day4/data/mindanao_ndvi_sample.csv`
- Reference: Medium - *Using PyTorch to Train an LSTM Forecasting Model*

Module 0: Setup & Context

Focus

- Ground the Mindanao drought challenge and toolkit
- Ensure Earth Engine + PyTorch environment is ready

Actions

- Quick NDVI refresher tied to agriculture risk
- Run `pip install earthengine-api` if needed; authenticate once per user
- Confirm notebooks open (GPU optional, CPU sufficient)

Module 1: Authenticate & Fetch NDVI

Focus

- Pull monthly Sentinel-2 NDVI for Bukidnon & South Cotabato straight from GEE

Actions

- Define province geometries (bounding boxes or GAUL features)
- Build GEE pipeline: filter S2 SR, compute NDVI, monthly median composites
- Export to pandas DataFrame and optionally cache to CSV for reuse

---

Module 2: Tidy & Build Sequences

Focus

- Prepare NDVI-only inputs for sequence modelling

Actions

- Inspect gaps/outliers; impute or drop empty months
- Create 12-month lookback / 1-month horizon windows per province
- Keep train/val/test split logic temporal (no shuffle, no leakage)

Module 3: Configure PyTorch LSTM

Focus

- Assemble the modelling pipeline around real NDVI data

Actions

- Wrap sequences in custom `Dataset`/`DataLoader`
- Define 2-layer LSTM (64/32 units, dropout 0.2) + dense head
- Select optimizer (Adam 1e-3), loss (MSE), metrics (MAE)

Module 4: Train & Monitor

Focus

- Fit the model while checking for convergence and overfitting

Actions

- Train ~70 epochs with validation monitoring
- Track train/val MSE + MAE; adjust learning rate/batch if unstable
- Discuss early stopping, reproducibility, and runtime tips

Module 5: Evaluate & Drive Decisions

Focus

- Convert forecasts into drought intelligence for stakeholders

Actions

- Invert scaling and plot actual vs predicted NDVI per province
- Compute MAE/RMSE, highlight months with largest residuals
- Apply NDVI < 0.40 threshold for alert table (precision/recall messaging)

Reflection & Wrap-up

- Key lessons: accessing satellite archives, disciplined validation, explainable alerts
- Capture open questions for Session 3 (XAI & emerging AI)
- Stretch: enrich features (rainfall/SPEI), expand to sequence-to-sequence or multi-region models