**I-GUIDE MODEL CARD**

The I-GUIDE Model Card is an easy-to-use tool that will allow you to create documentation for each model that you create or use in a project.

Using this tool will help facilitate transparency and reproducibility about your project. It will also help you comply with relevant policies of journals, funding agencies, and universities.

The Model Card applies to:

* Pre-existing models acquired from other sources, e.g., produced by other researchers;
* Models you and your collaborators produced yourselves;
* Models you and your collaborators produced by integrating two or more other models (e.g., coupling).

**Model Card Attribution**

This Model Card template is an adapted version of the I-GUIDE Data Card template, which itself is based on Google’s *Data Cards Playbook* (https://pair-code.github.io/datacardsplaybook/).  
It has been restructured to address key considerations for geospatial model transparency, performance evaluation, and ethical deployment, in alignment with the I-GUIDE research lifecycle.

This work is shared under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).  
License details: <https://creativecommons.org/licenses/by-sa/4.0/>

A logo with text on it

AI-generated content may be incorrect.

* + 1. **BASIC INFORMATION**

| Model Card ID Number | *MC-01* |
| --- | --- |
| Model Name | *XGBoost with Basic Features, Burn Histories and Embeddings* |
| Model Version | *1.0 (August 08, 2025)* |
| Persistent Identifier | [*https://doi.org/10.1145/2939672.2939785*](https://doi.org/10.1145/2939672.2939785) |
| Outputs Supported | [*Xie et al., (2019)*](https://doi.org/10.1007/s00521-018-3515-0)*;* [*Eskandari* *et al., (2020)*](http://link.springer.com/10.1007/s11069-020-04169-4)*;* [*Hanberry B. (2020)*](https://www.mdpi.com/2072-4292/12/18/2966) |
| Model Card Author | *Bikas Chandra Gupta, University of Texas at Arlington,* [*bikaschandra.gupta@uta.edu*](mailto:bikaschandra.gupta@uta.edu) |

* + 1. **MODEL OVERVIEW**

| Model Type | *(Select all that apply)*  ☑ AI model: Machine Learning (*Gradient Boosting)*  ☐ Statistical model: *(Specify type)*  ☐ Other: *(Specify)* | |
| --- | --- | --- |
| Purposes | *(Select all that apply)*  ☑ Classification  ☐ Decision support  ☑ Forecasting  ☑ Regression  ☐ Simulation  ☐ Spatial analysis  ☐ Other: *(Specify)* | |
| Domains of Application | *(Select all that apply)*  ☐ Climate science  ☐ Economics  ☑ Environmental impact modeling  ☑ Geospatial analysis  ☑ Hydrology  ☐ Population modeling  ☐ Other social systems modeling: *(Specify)*  ☐ Other: *(Specify)* | |
| Model Authors and Developers | *Bikas Chandra Gupta, University of Texas at Arlington,* [*bikaschandra.gupta@uta.edu*](mailto:bikaschandra.gupta@uta.edu) | |
| Source and Acquisition Method | ☐ Acquired (from external source)  ☑ Developed internally  ☐ Integrated from multiple models (e.g., coupled) | |
| User Licensing | ☑ Open source: *(XGBoost under Apache 2.0 License)*  ☐ Proprietary: *(Specify owner)*  ☐ Other restrictions on use: *(Specify restrictions)* | |
| Storage Location | ☑ Repository: [*https://github.com/ddominguez2293/I\_Guide\_Team\_2*](https://github.com/ddominguez2293/I_Guide_Team_2)  ☑ Project-specific storage: *Local project storage* |
| Access Control Policies | ☑ Open  ☐ Embargoed: *(Describe release timeline)*  ☐ Restricted: *(Describe access criteria)* |
| Use Case | *The model is designed to predict next-day wildfire burn progression at MODIS scale (500m) and explore fire risk mapping (susceptibility) in the Pantanal wetlands.* | |

* + 1. **MODEL INPUTS AND TRAINING DATA**

| Model Inputs | 1. *MODIS Burn Area (MCD64A1)* 2. *Lagged burn history (1–5 days)* 3. *ERA-5 Climate variables (temperature, precipitation, humidity, wind)* 4. *DEM (slope, elevation, aspect)* 5. *Satellite embeddings (Google Earth Engine Satellite Embeddings)* |
| --- | --- |
| Input Data Types | *(Select all that apply)*  ☑ Raster  ☑ Tabular  ☑ Time Series  ☐ Vector  ☐ Other: *(Specify)* |
| Training Data Used | 1. *MODIS Burned Area (MCD64A1.061,* [*https://doi.org/10.5067/MODIS/MCD64A1.061*](https://doi.org/10.5067/MODIS/MCD64A1.061)*)* 2. *ERA5 Climate reanalysis (*[*https://doi.org/10.24381/cds.68d2bb30*](https://doi.org/10.24381/cds.68d2bb30)*)* 3. *SRTM DEM (*[*https://doi.org/10.1029/2005RG000183*](https://doi.org/10.1029/2005RG000183)*)* 4. *Satellite embeddings (*[*https://doi.org/10.48550/arXiv.2507.22291*](https://doi.org/10.48550/arXiv.2507.22291)*)* 5. *Burn History (derived)* |
| Training Dataset Representativeness | * *Focused on the Pantanal wetlands; limited global generalizability.* * *Seasonal aggregation mitigates cloud gaps in Sentinel data.* * *Fire events concentrated in the dry season (June–Sept), so wet season fire representation is limited.* |

* + 1. **MODEL STRUCTURES**

| Feature Selection | * *Included climate variables, topography, land cover, burn lags, embeddings.* * *Excluded variables with high collinearity or missingness.* |
| --- | --- |
| Hyperparameters and Tuning | * *Learning rate (0.05–0.1)* * *Max depth (6–10)* * *n\_estimators (500–1000 with early stopping)* * *Subsample (0.8)* * *Colsample\_bytree (0.8)*   *(****Tuned via Bayesian Technique****)* |
| Software and Dependencies | *Python > 3.10; XGBoost 3.0.3; scikit-learn, pandas, numpy, geopandas, rasterio, xarray, rioxarray* |

* + 1. **MODEL PERFORMANCE AND EVALUATION**

| Validation Approach | ☐ Cross-validation  ☑ Holdout set  ☑ Time series split  ☐ Other: *(Specify)* |
| --- | --- |
| Evaluation Results | *Accuracy: 0.965*  *Precision: 0.964*  *Recall : 0.965*  *F1 Score : 0.965* |
| Testing or Validation Data Used | *MODIS Burned Area (MCD64A1.061,* [*https://doi.org/10.5067/MODIS/MCD64A1.061*](https://doi.org/10.5067/MODIS/MCD64A1.061)*)* |
| *(If model is integrated from multiple other models)*  Contribution of Constituent Models | *(Describe the contribution of each constituent model to the integrated model’s performance)* |

* + 1. **MODEL ADAPTATION AND CUSTOMIZATION (for acquired or integrated models only)**

| Source Models | *Not applicable – model developed from scratch using XGBoost library* |
| --- | --- |
| Availability of Source Model Code | *(Complete for each original model)*  ☐ Openly available: *(Include link)*  ☐ Restricted availability: *(Describe restrictions)*  ☐ Unavailable: *(Explain reason for unavailability)* |
| Modifications | *NA* |
| Training Data Adjustments | *NA* |

* + 1. **MODEL DEPLOYMENT AND USAGE**

| Computational Requirements | * *Runs on CPU* * *Typical training: ~1 hours on ~8-core CPU with 16 GB RAM* |
| --- | --- |
| Geospatial Considerations | * *Region-specific calibration (Pantanal)* * *Requires local climatology and land cover for transfer to other regions* |

* + 1. **TRANSPARENCY, EXPLAINABILITY, AND INTERPRETABILITY**

| Model Transparency | ☐ Fully transparent (rule-based, interpretable ML)  ☑ Partially transparent (some explainability features)  ☐ Black box (deep learning, complex ML models) |
| --- | --- |
| Explainability Features | ☑ Feature importance analysis  ☐ LIME  ☑ Sensitivity analysis  ☑ SHAP values  ☐ Other: *(Specify)* |
| Interpretability Challenges | *Complex interactions in boosted trees may reduce intuitive interpretability.* |
| Communication of Model Limitations | * *Model outputs include burn probability maps with associated uncertainty (via ensemble variance).* * *Limitations: only MODIS-scale (500m), does not capture fine-scale fire spread.* |

* + 1. **OTHER ETHICAL CONSIDERATIONS**

| Ethical Risks (Other Than Transparency, Explainability and Interpretability) | *(Select all that apply)*  ☑ Bias in training data: *(Pantanal-focused, may underperform elsewhere)*  ☑ Intentional misuse risks: *(misinterpreting as a real-time operational forecast)*  ☐ Privacy risks and surveillance: *(Specify)*  ☐ Security risks: *(Specify)*  ☐ Stigmatization of individuals or communities: *(Specify)*  ☐ Other: *(Specify)* |
| --- | --- |
| Measures Taken to Address Ethical Risks | * *Clear documentation of limitations* * *Recommendation for use only in research, not operational fire management* |
| Suitable Uses | * *Research on fire dynamics and susceptibility* * *Supporting environmental analysis of Pantanal and similar ecosystems* |
| Unsuitable Uses | * *Operational firefighting or real-time evacuation planning* * *Application outside Pantanal without retraining/validation* |