Frequently Asked Questions - Early Warning System RfP

Date uploaded: 9-Sept-2019
Last updated: 9-Sept-2019

Question	Answer
What is the status of the models developed by the Boston Consulting Group (BCG) and is more in depth material available to look at these models?	At this stage, only the Medium term forecasting model is developed (predicting 6 months into the future). This model is written in python and bash scripts and tested a variety of machine learning techniques. The coming two months (September/October, referred to as the *Prototype* phase) the EWS program will still continue to develop the current code and method further with the current tech partner. Thus, we cannot share the models in terms of scripts because these will be updated and improved during the prototype phase. The expected status (and deliverable) of the medium term forecasting model after the prototype phase with BCG is an automated, scalable method/product (expected Python 3) able to work with the current input datasets to generate predictions.
What Programming languages did BCG use during the design phase?	Python 3, Bash (will likely be replaced by Python 3)
What are the cloud services used by BCG?	Amazon Web Services (AWS) (S3 Bucket and EC2 instance)
What are the machine learning models tested?	Random Forest, Neural Network (not convolutional), Support Vector Machine and ensemble of these algorithms
Is the development of the reactive model based on Sentinel data out of the scope of this RFP?	The development of the reactive model is out of the scope of this RfP. The reactive model is however part of the EWS program.
There is mention of 1- year available funds, do you want a budget for 3 years' activities or of just 1 year?	We want a budget estimation for all deliverables mentioned in the RfP. Thus for the 3 years. The budget breakdown will be

	evaluated by deliverable. We allow for different cost options for the different deliverables.
What are the results achieved so far for the EWS program in terms of accuracy?	See image below - study area (50 km by 80 km)

Patch size: in RfP documentation referred to as 'hot-zone'.
Testing period: Dates of time span in which the accuracy assessment (test) is performed
AUC = Area Under Curve

Patch size	Test period start	Test period end	Confidence threshold	True positives	False positives	User's Accuracy	Accuracy	AUC
480 x 480 m	20171101	20180401	95%	4	6	40.00%	91.65%	50.11%
			75%	84	32	72.41%	91.94%	52.61%
			50%	597	278	68.23%	93.38%	68.40%
	20180501	20181001	95%	7	2	77.78%	90.46%	50.19%
			75%	72	14	83.72%	90.74%	51.97%
			50%	640	253	71.67%	92.51%	67.20%

Does the RfP applicant needs to provide information/data to discriminate legal from illegal deforestation in the EWS program under the RfP?	To determine the legality status, the EWS program depends on information to be supplied by the user board. How this information will be supplied is part of the ongoing engagement with the user board.
Is the current predictive model fully automated on data flow or does it require manual operations?	The predictive model should be fully automated by the end of the prototype phase. This includes predictive indicator generation, model training, and model evaluation.
How is the model documented? (Inline, documentation on exact data, data origin, data definitions, performed transformations, versions of libraries used, model definitions, set up test framework)	The EWS program has the majority of the documentation documented (in-line code, spreadsheets, slides). This information is aimed to be streamlined toward end of prototype phase.
Is data available to scale to Borneo/Sumatra?	Processed (and labelled) Sentinel 1 data on forest, deforestation, and forest degradation will be made available by the EWS program through the data provider (outside the RfP). Other data sources that are used in the design phase of the EWS program are expected to be available for scaling of the EWS program. As most of these data sets are public, open and free (spatial) datasets.
What are the current input datasets of the model?	The current data input consist of: - Satellite data (processed and labeled) - Elevation (public data) - Population density (public data) - Protected areas (public data) - Roads (public data)

What are the indictors of the predictive model?	Top 10 of the most important indicators for the current study area (Could change once EWS starts scaling to larger areas) - Distance to deforestation - Accumulated* deforestation/degradation - Distance to degradation - Distance to roads - Distance to population centers - % deforestation in 1 km radius - Patch density deforestation - Aggregation deforestation - Elevation - % not forest/deforestation/degradation/water in 1 km radius * Accumulated def. is a SarVision product that counts the number of times that deforestation/degradation are mapped against a baseline. Also includes water, forest, and not forest land cover classes.
What do we expect on extra data to become available?	Examples are (but not limited to): - Waterways - Logging roads - Canals - Palm oil mills - Forest type - Historic logging roads