

# Generation of space weather synthetic data based on machine learning approaches.

Felipe Meza-Obando<sup>1</sup>, Carolina Salas-Matamorros<sup>2</sup>, Juan Luis Crespo Marino<sup>1</sup>,  
Alexander Solis Lopez<sup>1</sup>

<sup>1</sup>Laboratorio de Inteligencia Artificial para las Ciencias Naturales (LIANA) Instituto Tecnológico de Costa Rica, <sup>2</sup>Centro de Investigaciones Espaciales (CINESPA) Universidad de Costa Rica

## Abstract

One of the major challenges when working with prediction tasks using space weather data, consist in the selection of the machine learning method to run such predictions. Most of the machine learning methods require an important amount of data, and depending on the database selected or generated, this is something not always easy to do.

In this research we are using the propagation speed ( $V_p$ ) of a coronal mass ejection, to predict the transit time (TT) and have a reliable value of arrival to Earth. The problem is that based on our sources of data and the time range selected for the study, the amount of data is limited.

To deal with this data situation, we are proposing the use We use machine learning based methods to generate synthetic data; methods such as SMOTE, ADASYN, Generative Adversarial Networks and Polynomial Regression. The main idea will be to generate synthetic data with each method based on different proposed models using just  $V_p$  and TT, and finally create an ensemble approach to consider a unique synthetic value, based on the ensemble results from the other methods.

By the increment of the database in size, we will be able in the next stages of this research to explore more machine learning options to run the prediction tasks, and obtain better accuracy results.

**Keywords:** space weather, machine learning, smote, gan.

## References

Gopalswamy, N., Lara, A., Yashiro, S., Kaiser, M. L., and Howard, R. A. (2001), Predicting the 1-AU arrival times of coronal mass ejections, *J. Geophys. Res.*, 106( A12), 29207– 29217, doi:[10.1029/2001JA000177](https://doi.org/10.1029/2001JA000177).

Srivastava, N., and Venkatakrisnan, P. (2004), Solar and interplanetary sources of major geomagnetic storms during 1996–2002, *J. Geophys. Res.*, 109, A10103, doi:[10.1029/2003JA010175](https://doi.org/10.1029/2003JA010175).

Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *J. Artif. Int. Res.* 16, 1 (January 2002), 321–357.

**NOTE: Email comments or for more details [fmeza@itcr.ac.cr](mailto:fmeza@itcr.ac.cr)**