Truth Data for Space Weather

Monica Ball, Andrew Jossi, Sienna Amorese, Will Biancarelli

PURDUE UNIVERSITY | SPACE ISAC

Overview & Background

Space weather impacts to satellites and spacecraft are a concern for the space industry and predicting the likelihood of space weather induced anomalies is valuable to satellite operators. Through analyzing space weather and known satellite anomalies, Data Mine students sought to develop a machine learning (ML) model which would ingest data on space weather conditions and predict the likelihood of anomalies. This capability would contribute to value offered by Space ISAC to its satellite operator members.

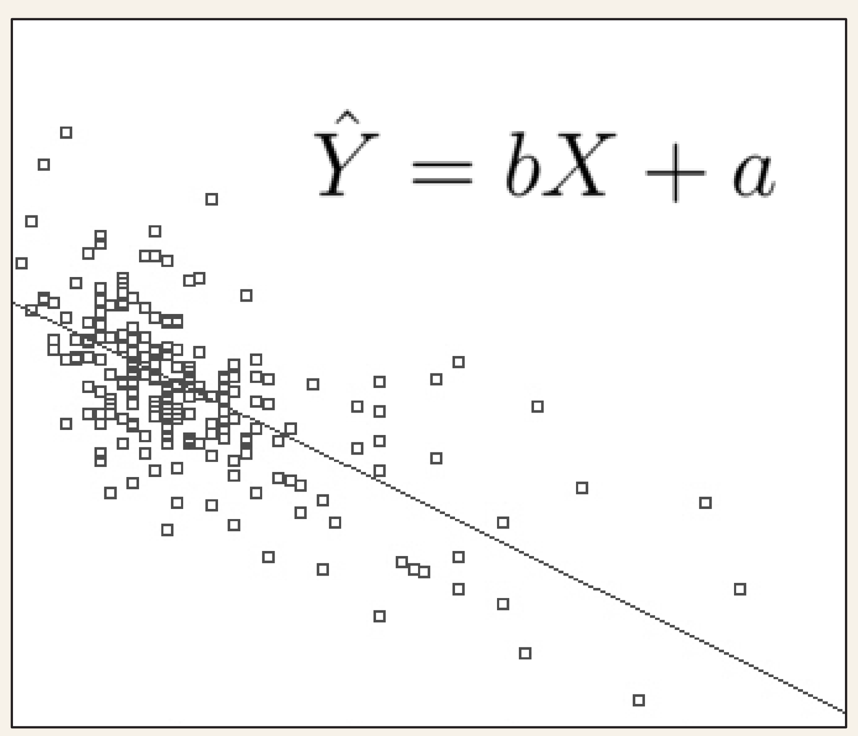
Machine Learning Model:

Baseline Model: Based on the structure of the data, there is a 50/50 chance that an anomaly will occur. This is due to the fact that half of the observations are anomalies and the other half is not. We aimed to improve on this baseline in order to predict anomalies in satellites.

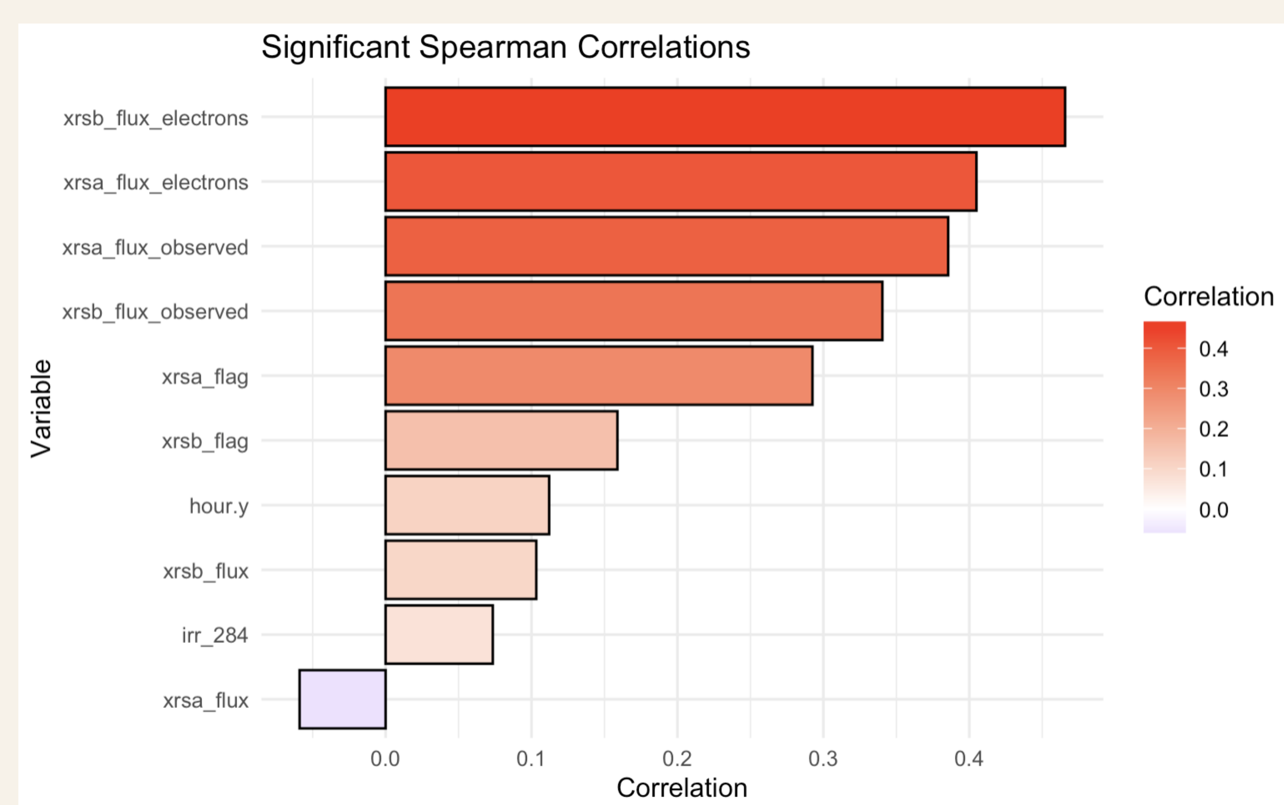
Data:

Data Collection: The team used open-source data from the National Oceanic and Atmospheric Administration's (NOAA) GOES-16 and GOES-17 weather satellites. These provided data on time of anomalies, and space weather. The team identified extreme ultraviolet radiation (EUV) and X-ray irradiance as good predictors of anomalies.

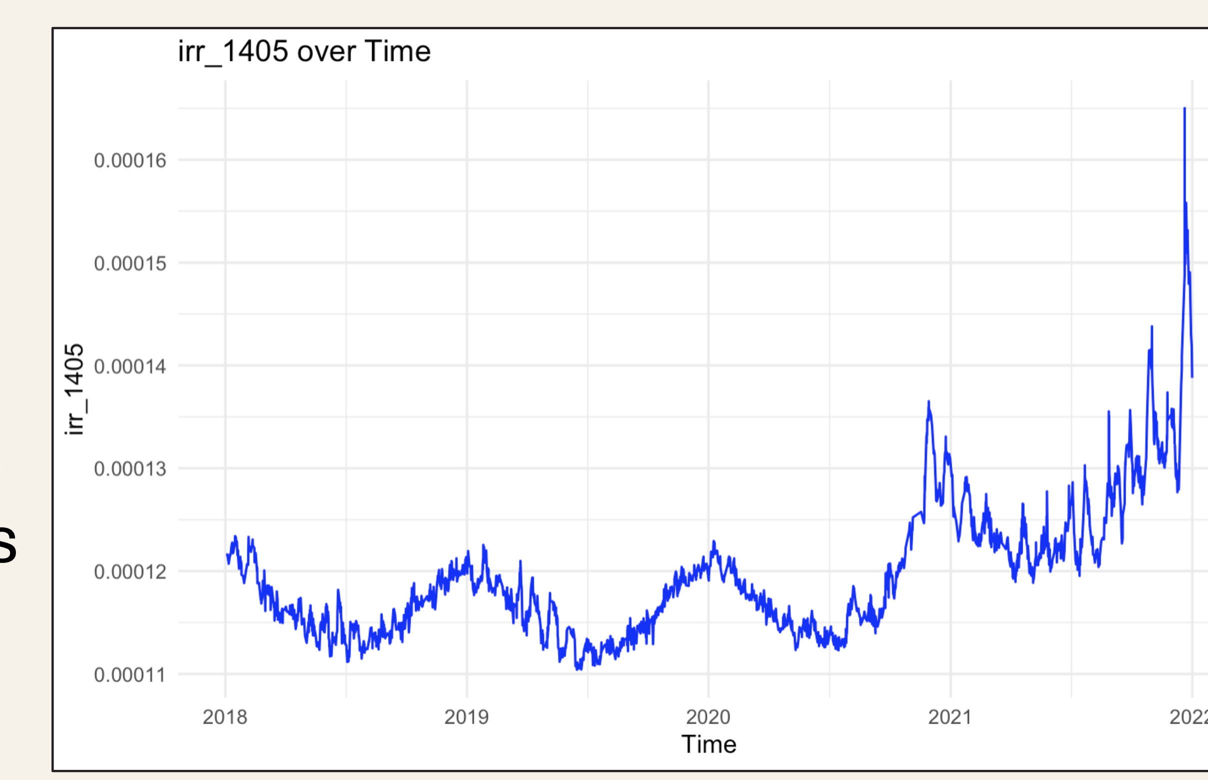
Data Processing: The team combined the data into a single data set. The data was then tailored by creating modeling variables. The team explored several predictor variables by plotting them against time and observed how they changed with time. This allowed the team to select important variables. Finally, the data was cleaned by eliminating non-useful variables and NA values. The data was split into 1,295 observations with anomalies and 1,295 without for a total of 2,590 observations.



Spearman Correlation: The team found the correlation between all variables and the response variable (anomaly = 1, no anomaly = 0). The most highly correlated variables were selected for the next model.



Model Tuning: The team switched to a logistic regression model, which is useful for predicting binary outcomes. The glm() function from R was used to predict anomalies with the most strongly correlated predictor variables.



Model Evaluation: The logistic regression model was found to have a cross validated accuracy of 72%. An area under the curve (AUC) of 0.79 shows a good ability to distinguish anomalies from non-anomalies. The model achieved a 77% sensitivity (correctly detecting anomalies) and 64% specificity (correctly identifying non-anomalies) which is important for reducing false alarms.

Discussion:

Limited open-source data relating space weather observations and observed satellite anomalies was a major limitation. The data used covered only geosynchronous orbit (GEO). The team would like to expand the project to incorporate low earth orbit (LEO), where many satellites are located. More complex deep learning models and a greater diversity of measurement types could improve upon the team's work so far. Future models could include: tree-based models for better handling of complex interactions; as well as deep learning models and support vector machines to capture non-linear, time series patterns.

Acknowledgments:

The team would like to thank: Hector Falcon, Samantha Dubois, Space ISAC for their guidance, as well as Jessica Jud, Ashley Arroyo, Dr. Dan Hirleman, and the Data Mine.