Project Progress Report

Vidyut Baskar

March 17, 2025

Space Weather Prediction using Machine Learning

Data Collection and Exploration:
Over the past several weeks, I have focused on gathering and exploring the historical space weather data necessary for building a predictive model. The primary dataset comes from NOAA's Space Weather Prediction Center (SWPC), providing timestamps of solar flares, solar wind measurements, and geomagnetic indices (e.g., Kp index). Data were collected through publicly available repositories and APIs. To prepare the dataset, I addressed missing values, aligned timestamps, and ensured consistent measurement units and time zones. Preliminary analysis shows a correlation coefficient of approximately **0.68** between solar wind speed and the Kp index, suggesting a meaningful relationship relevant to geomagnetic storm prediction.

Modeling and Preliminary Results:
I began with two baseline approaches:

- **ARIMA (Auto-Regressive Integrated Moving Average):** Provides a benchmark for univariate time-series forecasting. In initial tests, it achieved a Root Mean Squared Error (RMSE) of around **0.19** on the validation set.
- **Basic LSTM (Long Short-Term Memory):** A simple LSTM network trained on multiple correlated features (e.g., solar wind speed, IMF parameters, proton density). After minimal hyperparameter tuning, this model achieved an RMSE of **0.15**, indicating an improvement over the ARIMA baseline.

I also experimented with feature engineering by adding rolling averages and lagged variables, which slightly improved the LSTM's performance. These preliminary results suggest that deep learning models can capture complex temporal dynamics in space weather data more effectively than traditional statistical methods.

Next Steps:

1. **Hyperparameter Tuning:** Systematically refine the LSTM model's parameters (learning rate, hidden layers, sequence length) to further enhance prediction accuracy.
2. **Feature Refinement:** Incorporate additional derived features such as moving standard deviations, differences over time, and seasonality indicators.
3. **Transformer Models:** Explore Transformer-based architectures to potentially capture long-range dependencies more effectively than LSTMs.

4. **Expanded Evaluation:** Use classification metrics (e.g., precision, recall) to evaluate performance on forecasting "storm vs. non-storm" periods, in addition to regression metrics like RMSE and MAE.

Overall, the project is proceeding as planned. I have a functioning data pipeline, an initial demonstration of the value of deep learning models, and a clear roadmap for further experimentation and improvement. With continued refinement, the goal is to develop a robust and reliable system for predicting space weather events that may impact satellite operations and terrestrial infrastructure.