Aurora Predictions through Machine Learning Models

Skylar Pae | spae2@gmu.edu
CDS Department | George Mason University

RESEARCH QUESTION/ HYPOTHESIS

Can Space Weather Activity be Predicted by Machine Learning Models?

INTRODUCTION / CONTEXT

The purpose of this project is to utilize several data mining methods to predict space weather activity, specifically auroral activity. The methods proposed to predict auroral activity were Artificial Neural Networks (ANN), Bayesian Inference, Clustering, and Principal Component Analysis (PCA). While space weather phenomena such as aurora borealis and australis are somewhat unpredictable, NOAA records both solar and planetary data that allows scientists to make auroral predictions. Through the data provided by NOAA and pattern recognition, this model will provide more definitive answers to both scientists and aurora enthusiasts who wish to see or learn more about auroral activity.

DATA COLLECTION

The imported files contain Kp index values, solar magnetometer data, and solar plasma values (solar wind variables) over a 7-day period, spanning from the current day to 7 days prior. The Kp value is the most essential attribute in this dataset, as it is the average measurement of global magnetic activity ("Tips on Viewing the Aurora"). This index ranges from 0 to 9, where higher values indicate greater amounts of activity. Solar plasma values indicate the abundance of protons emitted from the sun, while magnetometer values represent the strength of the magnetic field carrying the plasma ("Science: Solar Wind"). The code imports the JSON file directly from NOAA's product website using a URL. Because NOAA updates this information daily, this provided the opportunity to import data per the current date to output the daily aurora prediction for the user.

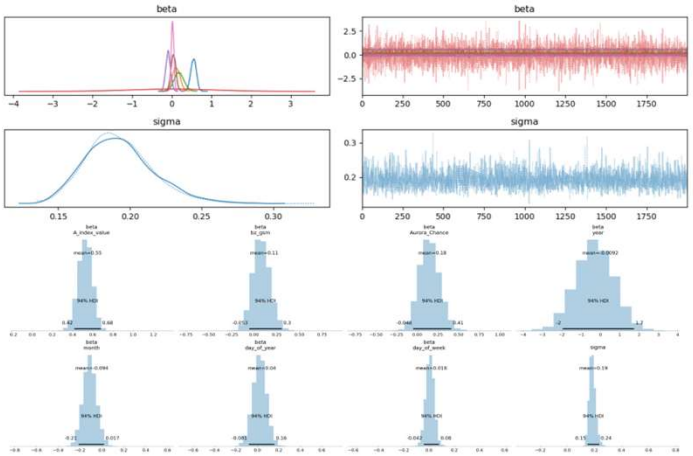

Figure 1: Bayesian Inference - Convergence & Distribution of Parameters (7/5/25)

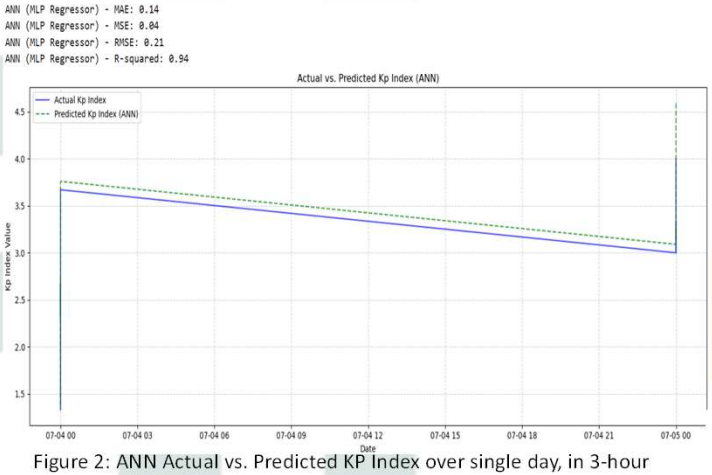

Figure 2: ANN Actual vs. Predicted KP Index over single day, in 3-hour intervals (7/5/25)

DATA MODELING / METHODOLOGY

All proposed methods of modeling (ANN, Bayesian Inference, Clustering, and PCA) were applied to the historical geomagnetic datasets extracted from NOAA. PCA was performed to highlight variability between attributes, Clustering to identify correlations between aurora predictor variables, ANN to model the complex patterns of the variables, and Bayesian Inference to determine probabilistic outcomes for the aurora predictions.

EVALUATION / RESULTS

After applying these machine learning models to the NOAA 7-day historic data, Bayesian Inference and ANN proved to be the most effective in reporting aurora predictions through Kp index values. The Bayesian inference's success was to be expected, as it provides probabilistic outcomes and aligns perfectly with predicting auroral activity. Through trace and posterior plots, shown in Figure 1, the results indicate successful convergence of parameters. By plotting an uncertainty interval, as shown in Figure 3, the results demonstrate that the Bayesian Inference model was able to predict values with more than 50% accuracy. Through R-squared calculation, the accuracy of the ANN method was determined. As shown in Figure 2, the ANN had an R-squared value of 0.94, indicating that the model could explain 94% of the variation in the prediction data.

CONCLUSION

The Bayesian inference and ANN models, with probabilistic determination and variable pattern recognition, were the most effective methods for predicting auroral activity through the Kp index values. Bayesian Inference accurately predicted Kp index values by more than 50%, and ANN by 94%. By using Bayesian Inference and ANN machine learning models on current-date NOAA Kp index readings, atmospheric scientists and aurora enthusiasts alike can find definitive daily predictions to plan their aurora-related adventures accordingly.

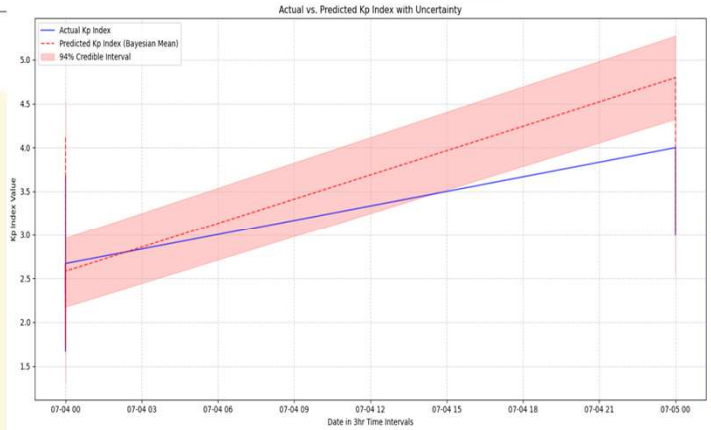

Figure 3: Bayesian Inference - Actual vs. Predicted Kp Index with Uncertainty/Credibility Interval (7/5/25)

REFERENCES:

"Science: Solar Wind", *NASA - Jet Propulsion Laboratory* https://www.jpl.nasa.gov/nmp/st5/SCIENCE/solarwind.html#:~:text=The%20solar%20wind%20is%20created,lines%20that%20extend%20radially%20outward. Retrieved 5 July 2025.
"Tips on Viewing the Aurora", *NOAA - Space Weather Prediction Center*, https://www.swpc.noaa.gov/content/tips-viewing-aurora. Retrieved 5 July 2025.
"Index of Products", *NOAA - Space Weather Prediction Center*, https://services.swpc.noaa.gov/products/. Retrieved 5 July 2025.

ACADEMIC INTEGRITY STATEMENT

I used AI to help debug and help structure my code in steps. AI also helped with converting the URL/JSON format to CSV, as I hadn't learned how to previously.
See transcript here: https://g.co/gemini/share/105b9d6be3e2