

Solar Flare Prediction Project Report

Goal:

To use machine learning to predict strong solar flares.

Data:

We used a dataset of solar activity with features like:

- Sunspot size
- Spot distribution
- Magnetic activity
- Flare history

Models Tested:

1. Ridge Classifier
2. K-Nearest Neighbors (KNN)
3. Random Forest

Best Model:

Random Forest gave the best results with high accuracy, precision, and recall.

What We Learned:

- Machine learning can help predict dangerous solar flares.
- It is important to clean and transform the data before training the model.
- Random Forest handles complex data well.

Next Steps:

- Try more data from NASA or NOAA.
- Use deep learning models.

- Build a web app to make real-time predictions.

Conclusion:

This project used machine learning to predict severe solar flares based on solar activity data. Among the models tested, Random Forest achieved the best performance with a weighted F1-score of 0.973, outperforming Ridge (0.971) and KNN (0.969).

All models demonstrated strong precision and recall, confirming their ability to handle class imbalance effectively. Analysis also revealed that Zurich classes F and H were highly associated with severe flares, making them strong predictive features.

The preprocessing pipeline combining encoding and scaling proved effective and forms a solid foundation for future solar flare forecasting tools and real-time applications.