Advanced Geomagnetic Storm Forecasting Using DSCOVR and CASSIOPE Missions: Developed for SPACE APPS Challenge 2023

Jorge Lozano* Melissa Amado†

June 27, 2024

Abstract

Geomagnetic storms pose significant risks to modern technologies, particularly GPS satellite systems and electrical power grids. These storms result from solar wind and storms from the Sun that affect Earth's magnetic field. Accurate prediction of these storms is crucial but challenging due to the variable travel time of solar material to Earth. The Deep Space Climate Observatory (DSCOVR) provides critical data for forecasting geomagnetic storms by measuring solar wind parameters. However, with its mission extended beyond the initial five years and instruments experiencing sensitivity issues, there is an urgent need for improved prediction models.

This project aims to develop an integrated database combining data from DSCOVR and CASSIOPE missions to predict the planetary K-index (Kp) with a lead time of 20 minutes to 1 hour. The integrated database will serve as input for advanced predictive models of geomagnetic activity. The methodology involves collecting data from 2016 to 2023, preprocessing to clean and normalize the data, temporal synchronization of timestamps, and data integration. A neural network model is implemented to classify Kp values using historical data, with the aim of providing early warnings of geomagnetic storms.

The results include a detailed correlation analysis between various solar wind features and the Kp index, identifying key predictors for geomagnetic activity. The machine learning model shows good overall performance with an accuracy of 82%, but specific areas need further optimization. This integrated approach enhances the predictive capabilities for geomagnetic storms, contributing to better preparedness and mitigation of their impacts on modern technology.

*Email: jorge.fernando.lozano@gmail.com, BeE3LabTech, Universidad Nacional de Ingeniería
†Email: melissa.amadovidal@gmail.com, BeE3LabTech, Universidad de Salamanca

1 Background

Geomagnetic storms on Earth are a menace to many modern technologies, particularly GPS satellite systems and electrical power grids. These storms occur when strong gusts of wind or storms from the Sun traverse interplanetary space and reach Earth, deforming Earth's magnetic field and showering particles into Earth's magnetic poles. These storms are notoriously difficult to predict. Even when solar flares and eruptions are observed that may cause a geomagnetic storm, the travel time for material to reach Earth could be anywhere from about two to four days (or it could miss Earth entirely).

NOAA's space weather station, the Deep Space Climate Observatory (DSCOVR), orbits about a million miles from Earth in a unique location called Lagrange point 1, which basically allows it to hover between the Sun and our planet. From that vantage point, DSCOVR measures the plasma that may cause geomagnetic storms hours before it reaches us– ideally providing an early warning of what's coming our way. The time that it takes for that plasma to reach Earth and trigger a geomagnetic storm might be anywhere from about 15 minutes to a few hours.

NOAA uses measurements of the solar wind density, temperature, speed, and magnetic field to run computer simulations of the Earth's magnetic field and atmosphere. Based on those simulations, NOAA forecasts when a geomagnetic storm will occur and how strong it will be. The strength of the geomagnetic storm is measured on a scale called the Planetary K-index (Kp).

The DSCOVR mission, which was initially planned for five years, is now in its eighth year. Although the instrument onboard DSCOVR that measures the solar wind's magnetic field continues to function very well, the instrument that measures the solar wind density, temperature, and speed has lost sensitivity and experiences faults and anomalies from time to time. These faults are unpredictable and can even be difficult to catch in real time, making them particularly troublesome when they occur during events that may cause storms. To make matters more urgent, the Sun is nearing the peak activity phase of its 11-year cycle like show in figure 1; these storms are more frequent now than at any point in DSCOVR's mission.

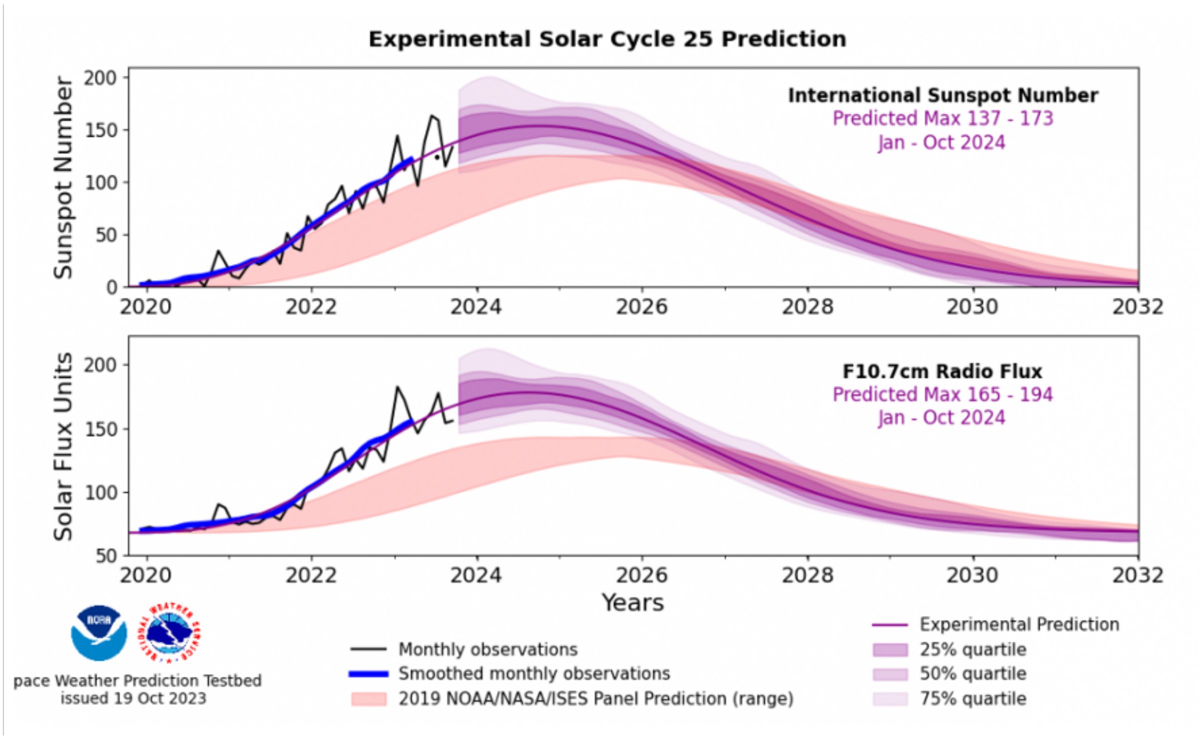

Figure 1: Solar Cycle

2 Problematic Reality

Space weather, particularly solar storms, is influenced by the Sun's 11-year cycle. During the peaks of this cycle, solar activity increases significantly, leading to a series of consequences in our technological and natural environment:

1. **Communication Interference**: Intense solar emissions can cause distortions in communication signals, affecting both terrestrial networks and satellite communications.

2. **Impacts on Satellites and Space Equipment**: Charged particles released during a solar storm can damage satellites and other space equipment, reducing their lifespan and compromising their functionality.

3. **Power Grid Disruptions**: Geomagnetic currents induced by these storms can overload power grids, leading to power outages and damage to critical infrastructure.

3 Literature Planetary K-index (Kp)

The Planetary K-index, commonly referred to as Kp, is a global geomagnetic activity index that quantifies disturbances in the Earth's magnetic field. It is derived from the K-index, which measures the fluctuations in the horizontal component of the magnetic field at specific geomagnetic observatories. The Kp index is an averaged measure taken from multiple observatories worldwide, making it a standardized and comprehensive indicator of geomagnetic activity.

3.1 Significance of the Kp Index

The Kp index is crucial for monitoring and predicting geomagnetic storms. These storms can have significant impacts on various technological systems, including satellite operations, GPS navigation, radio communications, and power grids. High Kp values indicate strong geomagnetic activity, which is typically associated with geomagnetic storms.

3.2 Scale and Interpretation

The Kp index ranges from 0 to 9, where each integer step represents an order of magnitude increase in geomagnetic activity.

1. **Kp = 0**: Very quiet geomagnetic conditions.
2. **Kp = 1 to 3**: Quiet to unsettled conditions, with minimal impact on technology.
3. **Kp = 4**: Active geomagnetic conditions, indicating minor geomagnetic storms.
4. **Kp = 5**: Minor geomagnetic storm conditions, which can cause minor disruptions in satellite and communication systems.
5. **Kp = 6**: Moderate geomagnetic storm conditions, with potential for more significant impacts on technological systems.
6. **Kp = 7**: Strong geomagnetic storm conditions, which can lead to disruptions in power grids and navigation systems.
7. **Kp = 8**: Severe geomagnetic storm conditions, causing widespread technological disruptions.
8. **Kp = 9**: Extreme geomagnetic storm conditions, leading to severe and widespread disruptions in various systems

3.3 Applications

1. **Space Weather Forecasting**: The Kp index is used by space weather forecasters to predict geomagnetic storm activity and issue warnings to affected industries and sectors.
2. **Technological Impact Mitigation**: By monitoring Kp values, operators of satellites, power grids, and communication systems can take preventive measures to mitigate the impacts of geomagnetic storms.

3. **Scientific Research**: The Kp index is a valuable tool for scientists studying the interactions between the solar wind and Earth's magnetosphere, as well as the broader effects of space weather on our planet.

In summary, the Kp index is a vital metric for understanding and predicting geomagnetic activity, helping to protect and manage the technological infrastructure that modern society relies upon. Its standardized measurement and wide-ranging applications make it an essential component of space weather monitoring and research.

4 Objetives

1. **Develop an Integrated Database**: Create a comprehensive database that combines real-time and historical data from the DSCOVR and CASSIOPE missions. This database will include measurements of solar wind density, temperature, speed, and magnetic field, along with ionospheric data, to provide a robust dataset for analysis.

2. **Predict the Planetary K-index (Kp)**: Utilize the integrated database to develop predictive models capable of forecasting the Kp index with a lead time of 20 minutes to 1 hour. This will enhance early warning systems for geomagnetic storms, allowing better preparedness and mitigation strategies.

3. **Implement a Classification Model**: Build and evaluate a neural network-based classification model using the integrated data to accurately predict Kp values. The model should achieve high accuracy and reliability, demonstrating its effectiveness through rigorous validation techniques such as K-Fold cross-validation.

4. **Conduct Correlation Analysis**: Perform a detailed correlation analysis between the variables measured by DSCOVR and CASSIOPE and the Kp index. Identify key predictors that have a significant impact on geomagnetic activity, thereby improving the model's predictive power.

5. **Optimize Predictive Accuracy** Continuously refine and optimize the predictive model by tuning hyperparameters and employing advanced machine learning techniques. Aim to achieve an overall model accuracy of at least 82%, with specific improvements in the classification of higher Kp values.

6. **Enhance Data Processing Capabilities**: Develop and utilize robust data processing methods to clean, normalize, and synchronize the datasets. Ensure that the integrated database is of high quality, facilitating accurate analysis and reliable model training.

5 Data Sources

In this chapter, we delve into the critical data sources used for predicting geomagnetic storms. The integration of data from the CASSIOPE and DSCOVR missions forms the backbone of our predictive model. By leveraging the unique capabilities and measurements of these two missions, we can create a robust and comprehensive dataset. This dataset not only facilitates the prediction of the planetary K-index (Kp) but also enhances our understanding of space weather phenomena. The following subchapters provide detailed descriptions of each mission, including their objectives, instruments, and scientific significance, as well as the methodology employed to integrate and process their data.

5.1 CASSIOPE Mission

CASSIOPE (CAscade, Smallsat and IOnospheric Polar Explorer) is a Canadian space mission that show in figure 2 launched on September 29, 2013, by the Canadian Space Agency (CSA). The mission has two main objectives: to conduct scientific research on the ionosphere and space weather, and to test data communication technologies.

Figure 2: Cassiope Mission

5.1.1 The main Components

1. **e-POP (Enhanced Polar Outflow Probe)**:

   (a) **Objective**: To study the interaction between the solar wind and Earth's magnetosphere, focusing on the flow of ions in polar regions and the effects of geomagnetic storms.

   (b) **Instruments**:

      i. **IRIS (Imaging and Rapid-scanning Ion Mass Spectrometer)**: Measures the composition and energy of ions in the ionosphere.

      ii. **RRI (Radio Receiver Instrument)**: Studies natural and artificial radio waves in the ionosphere.

      iii. **MGF (Magnetic Field Instrument)**: Measures Earth's magnetic field.

      iv. **NMS (Neutral Mass Spectrometer)**: Measures the density and composition of neutral particles.

      v. **CER (Charging and Relaxation Instrument)**: Measures the charging and discharging of particles in the ionosphere.

2. **Cascade**:

   (a) **Objective**: To test satellite communication technologies and demonstrate the feasibility of a high-capacity data transmission system from space.

   (b) **Instruments**: Communication equipment to test and validate data transmission technologies.

5.1.2 Orbit

1. **Type of Orbit**: Elliptical polar orbit.

2. **Orbital Altitude**: The altitude of CASSIOPE's orbit ranges from approximately 325 km to 1,500 km above Earth's surface. This orbit allows the instruments on board the satellite to conduct detailed observations of polar regions and the ionosphere.

5.1.3 Scientific and Technological Importance

1. **Scientific Research**:

   (a) **Space Weather**: CASSIOPE provides critical data on space weather, which affects various technologies on Earth, including satellites, communications, and power grids.

(b) **Ionosphere**: The mission studies the dynamics of the ionosphere and its interactions with the solar wind, enhancing our understanding of these phenomena and their impact on Earth.

2. **Communication Technologies**:Cascade, the communication payload, aims to demonstrate and validate new data transmission technologies, which can lead to significant improvements in satellite communications.

5.2 DSCOVR Mission

DSCOVR (Deep Space Climate Observatory) that show at figure 3 is a mission launched on February 11, 2015, by NASA, NOAA (National Oceanic and Atmospheric Administration), and the U.S. Air Force. The mission's primary objective is to monitor space weather and provide early warnings of geomagnetic storms, as well as to observe Earth's climate from a unique vantage point.

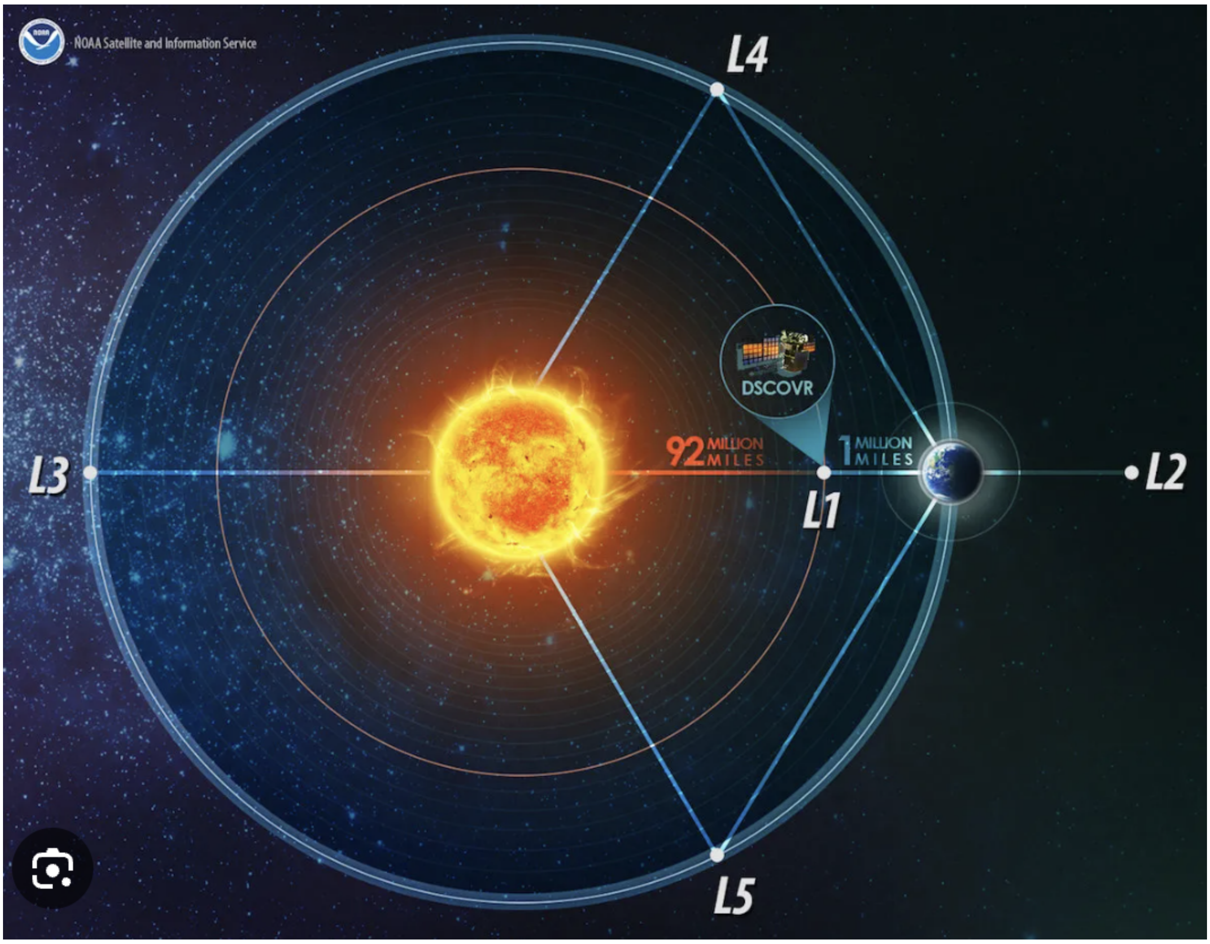

Figure 3: DSCOVR Mission

5.2.1 Main Components

1. **PlasMag (Plasma-Magnetometer) Suite**:
   (a) **Objective**: To measure the solar wind and its magnetic field, which are crucial for predicting space weather events like geomagnetic storms.
   (b) **Instruments**:
      i. **Faraday Cup**: Measures the density, temperature, and speed of the solar wind.
      ii. **Magnetometer (MAG)**: Measures the intensity and direction of the interplanetary magnetic field.
2. **EPIC (Earth Polychromatic Imaging Camera)**:
   (a) **Objective**: To observe and monitor Earth's atmosphere and surface, capturing images of the entire sunlit side of Earth.
   (b) **Instruments**: A camera that provides daily images of Earth, which are used for various climate and environmental studies.
3. **NISTAR (National Institute of Standards and Technology Advanced Radiometer)**:
   (a) **Objective**: To measure the Earth's radiation budget, including the amount of solar energy reflected and emitted by Earth.
   (b) **Instruments**: A radiometer that helps in understanding the Earth's energy balance and its impact on climate change.

5.2.2 Orbit

1. **Type of Orbit: Lagrange point 1 (L1)**
2. **Orbital Location**: Approximately 1 million miles (1.5 million kilometers) from Earth, between the Earth and the Sun. This position allows DSCOVR to maintain a constant view of the sunlit side of Earth and monitor the solar wind.

5.2.3 Scientific and Technological Importance

1. **Space Weather Monitoring**:
   (a) **Early Warnings**: DSCOVR provides real-time data on solar wind and magnetic field conditions, enabling NOAA to issue early warnings of geomagnetic storms that can affect satellites, power grids, and communication systems on Earth.
   (b) **Data Utilization**: The data collected are crucial for understanding the dynamics of space weather and mitigating its impacts on modern technology.

2. **Climate Observation**:
   (a) **Earth Monitoring**: EPIC and NISTAR provide valuable data for monitoring Earth's climate, including atmospheric conditions, cloud cover, and radiation balance.
   (b) **Climate Change Research**: The continuous observation of Earth helps in tracking changes in climate and supports various environmental studies.

6 Methodology for Creating an Integrated Database

In this chapter, we outline the comprehensive methodology employed to create an integrated database from the DSCOVR and CASSIOPE missions. The aim is to harness the strengths of both datasets to improve the prediction accuracy of geomagnetic storms. This chapter covers the detailed steps involved in data collection, preprocessing, temporal synchronization, and integration.

6.1 Data Collection

For this project, real-time or near-real-time data is collected from 2016 to 2023. DSCOVR data includes measurements of solar wind density, temperature, speed, and interplanetary magnetic field, obtained through the Faraday Cup and Magnetometer (MAG). On the other hand, CASSIOPE data comes from the e-POP instruments, such as IRIS, RRI, MGF, NMS, and CER, covering variables like ion composition and energy, radio waves, Earth's magnetic field, neutral particle density and composition, and particle charging and discharging.

6.2 Data Preprocessing

Data preprocessing involves cleaning and normalizing the collected data. Outliers, anomalies, and missing data are identified and removed using interpolation techniques to fill gaps and smooth time series. Additionally, variables are scaled to a common range using z-score or min-max normalization techniques, facilitating integration and comparative analysis.

6.3 Temporal Synchronization

To ensure data comparability, temporal alignment of timestamps from both data sources is performed. This is achieved using temporal interpolation techniques to align data with different sampling frequencies. Subsequently, a unified time series with consistent time intervals for both databases is generated.

6.4 Data Integration

Data integration involves combining corresponding variables from DSCOVR and CASSIOPE into a single data structure. A data schema is created to allow efficient access and analysis of all integrated variables. Additionally, labels and metadata are added to each record to identify the data source and measurement conditions.

6.5 Correlation Analysis

Correlation analysis focuses on calculating the relationship between variables measured by DSCOVR and CASSIOPE with the Kp index values estimated by CASSIOPE. This helps identify significant patterns and relationships between variables and Kp values, which is crucial for developing accurate predictive models.

6.6 Implementation of a Predictive Model

A neural network based on Sklearn has been implemented to create a classification model that predicts the Kp value on Earth with a 20-minute anticipation in the areas where CASSIOPE operates, based on DSCOVR data. This model uses historical integrated data from both missions to train the neural network, evaluating its performance and adjusting hyperparameters to optimize prediction accuracy and speed.

7 Implementation

In this chapter, we detail the practical implementation of the methodologies discussed in the previous chapter. This includes the generation and processing of the dataset, the application of machine learning techniques, and the creation of a neural network model for predicting geomagnetic storms. The steps outlined here demonstrate how the integrated data from DSCOVR and CASSIOPE is utilized to train and validate the predictive model. The implementation process ensures that the model is robust, accurate, and capable of providing early warnings of geomagnetic activity. Through rigorous data processing and model evaluation, we aim to achieve a high level of prediction accuracy and reliability.

7.1 Dataset Generation

The database class that shows in figure 4 facilitates the integration, analysis, and visualization of data from multiple sources. It allows for loading CSV and NetCDF files, converting and synchronizing dates, and merging datasets based on specific time intervals. The class includes methods for generating combined datasets, plotting time series of various measurements, and managing data using pandas and matplotlib. With these capabilities, database class optimizes data processing and analysis for projects that require the integration of different databases and the visualization of results.

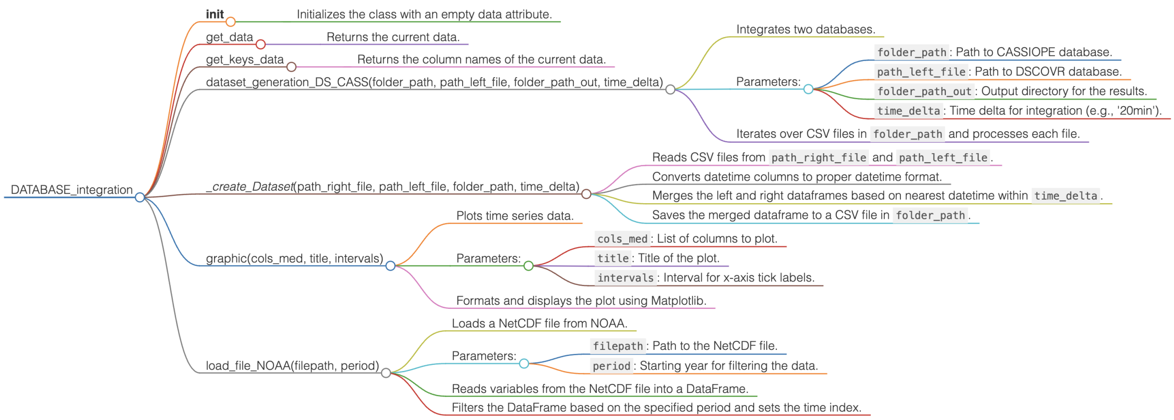

Figure 4: Dataset Generation Class

7.2 Dataset Proccessing

The dataset class that shows in figure 5 provides a comprehensive suite of methods for preparing, analyzing, and visualizing datasets in Python. It includes functionality to set indexes, shuffle rows, drop specified columns, and handle missing data. The class can calculate and visualize correlation matrices, clean outliers from data, and perform detailed attribute analysis by generating histograms. Additionally, it supports loading data from CSV files and dropping columns with high null value counts. With these capabilities, the dataset class facilitates efficient data preprocessing and exploratory data analysis, making it a valuable tool for machine learning and statistical tasks.

Figure 5: Dataset Proccessing Class

7.3 Machine Learning Process

The NeuronalNetwork file defines the RedNeuronal class shows at figure 6, which facilitates the creation, evaluation, and hyperparameter tuning of neural network models using ‘scikit-learn‘. The class includes methods for evaluating models through cross-validation and on complete datasets, generating neural models using K-Fold cross-validation, making predictions, and tuning model hyperparameters using GridSearchCV or RandomizedSearchCV. It also provides functions to identify the best hyperparameters and activation functions, scale data, and manage trained models. In summary, ‘RedNeuronal‘ is a comprehensive tool for efficiently building and optimizing neural networks.

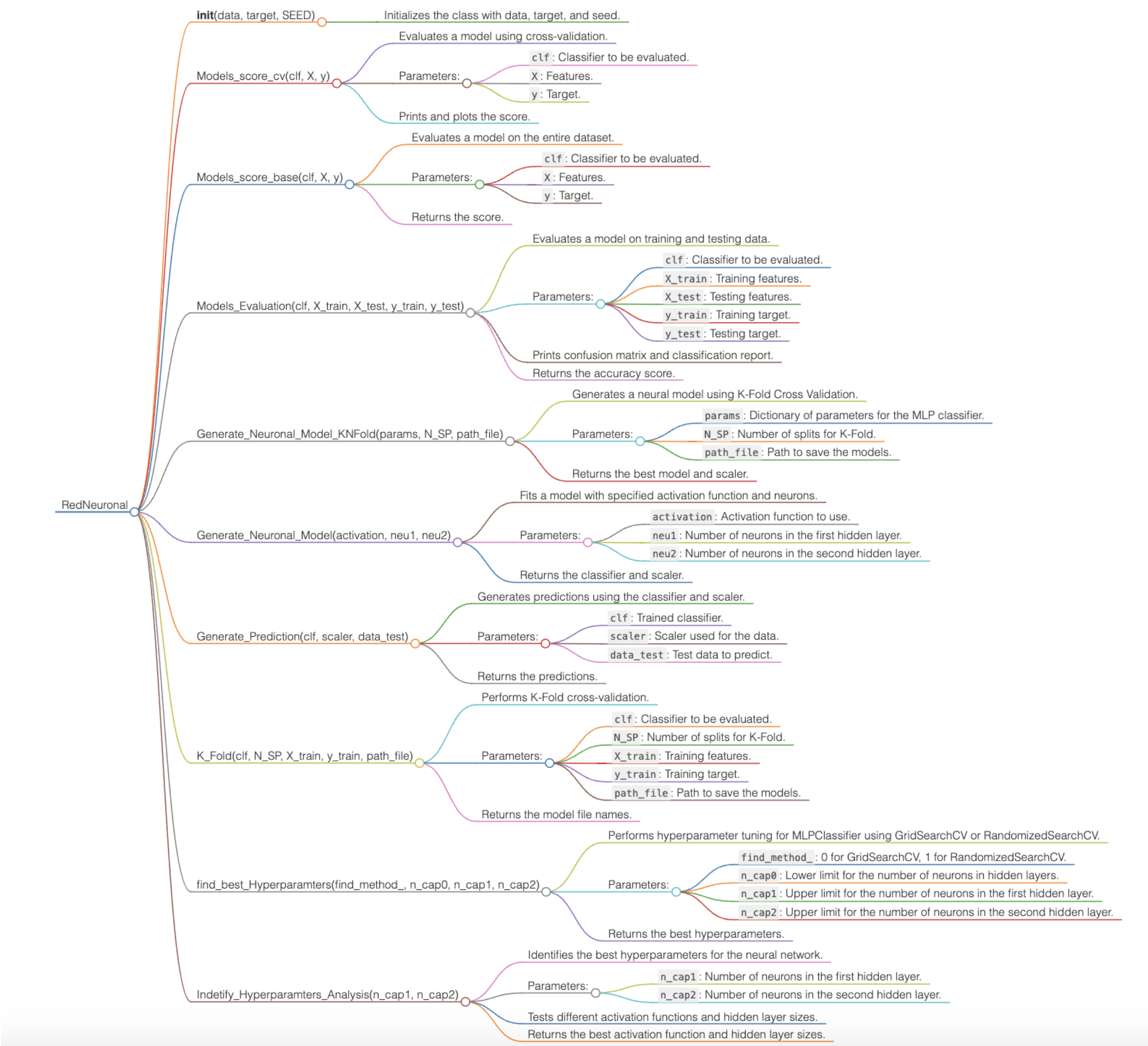

Figure 6: Machine Learning Proccess Class

8 Results

In this chapter, we present the results obtained from implementing the predictive model for geomagnetic storms. This includes a detailed analysis of the dataset, the correlations identified between various solar wind features and the planetary K-index (Kp), and the performance of the machine learning model. The results highlight the effectiveness of the integrated dataset and the neural network model in predicting Kp values with high accuracy. Additionally, we discuss the implications of these findings for space weather forecasting and the potential improvements that can be made to further enhance prediction accuracy. The following sections provide a comprehensive breakdown of the descriptive analysis, correlation analysis, and machine learning model results.

8.1 Descriptive Analysis

The figure 7 illustrates the distribution of various numerical variables from the dataset collected for the Solar Storms Prediction NASA Challenge. Each subplot represents the distribution of a specific variable, showcasing its frequency and density over different values. The data has been sourced from multiple missions, including DSCOVR and CASSIOPE, and includes key parameters related to solar wind and geomagnetic activity.

This comprehensive visualization provides valuable insights into the underlying patterns and behaviors of these variables, which are crucial for understanding and predicting geomagnetic storms. By examining the histograms and kernel density estimates (KDE) for each variable, we can identify the range, central tendency, and variability, as well as detect any potential outliers or anomalies. This foundational analysis is essential for further exploration and model development aimed at improving the accuracy of geomagnetic storm forecasts.

Key variables depicted in the graph include components of the magnetic field (`bx_gse`, `by_gse`, `bz_gse`), various solar wind features (denoted as fs), the total magnetic field magnitude (bt), and the planetary K-index (Kp), which measures the severity of geomagnetic storms. The distributions of these variables provide a detailed overview of the data's characteristics, setting the stage for advanced predictive modeling and analysis.

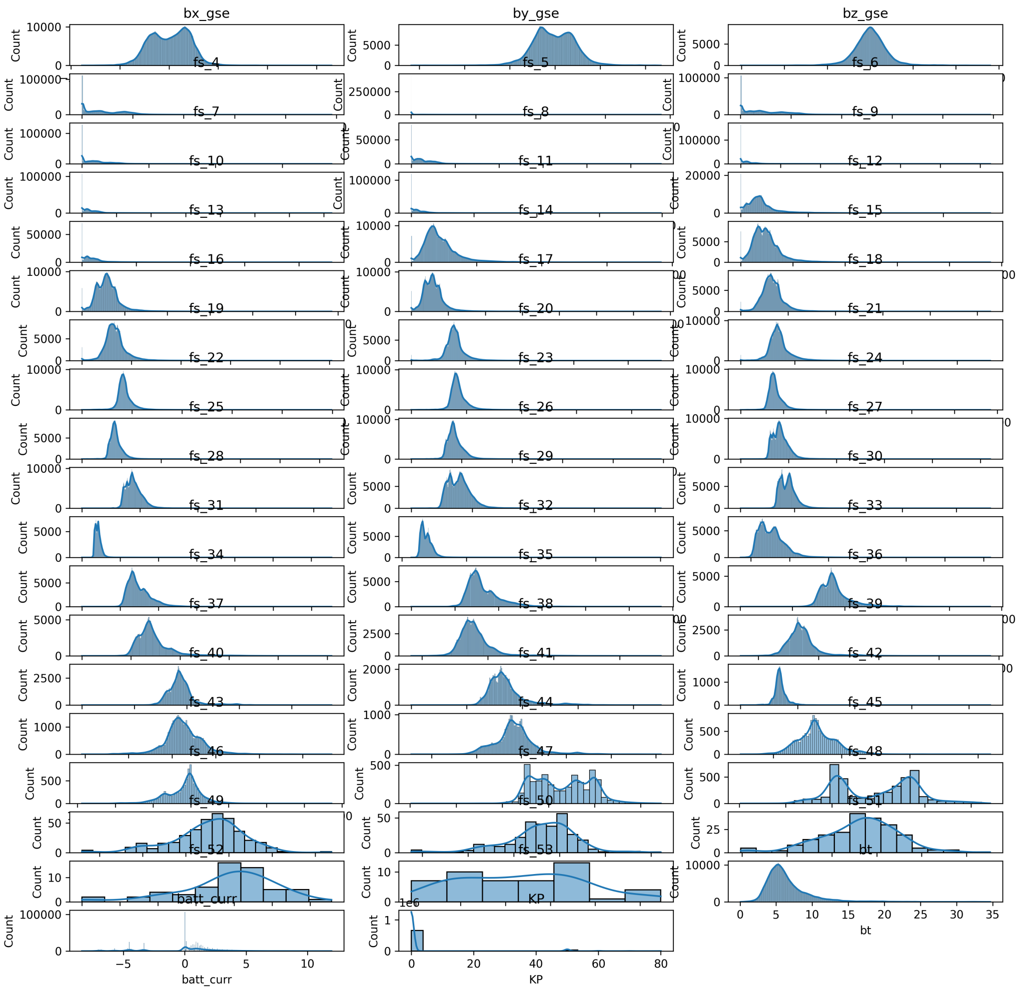

Figure 7: Density

Also, The figure 8 illustrates the distribution of the ‘bt‘ variable, representing the total magnetic field magnitude, across different levels of the planetary K-index (Kp). The Kp index values, indicated by different colors in the legend, range from 0 to 80, reflecting various levels of geomagnetic storm activity.

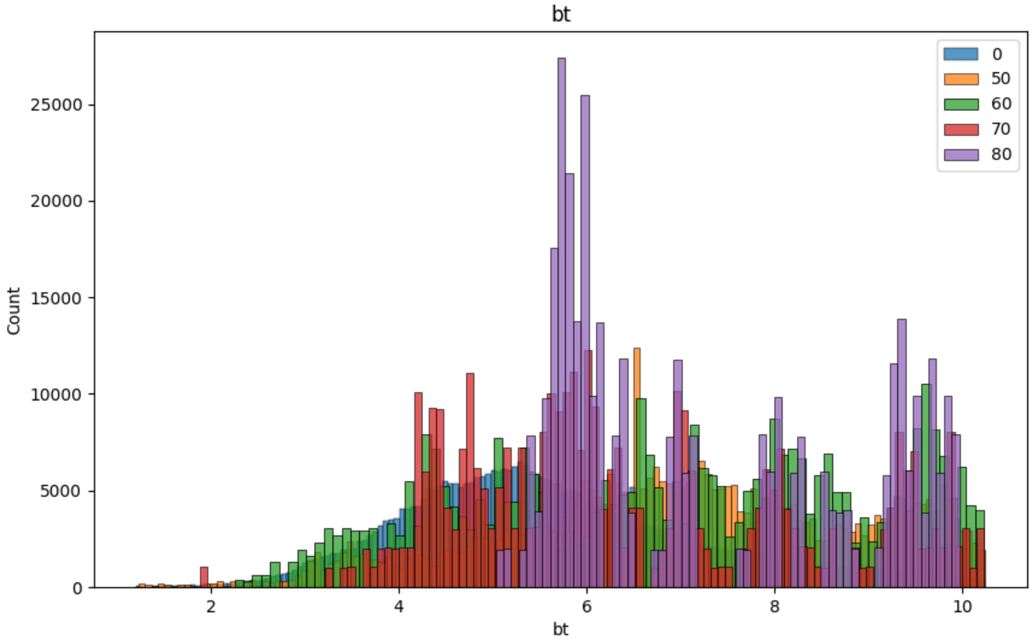

Figure 8: bt vs KP

Each bar in the histogram represents the frequency (count) of 'bt' values within specific bins, providing a clear visual representation of how the magnetic field magnitude is distributed for each Kp index category. The following observations can be made from the graph:

8.2 Correlations Analysis

The figure 9 presents a correlation matrix illustrating the relationships between various variables measured by the instruments of the DSCOVR and CASSIOPE missions. The included variables are components of the magnetic field in geocentric solar ecliptic (GSE) coordinates, several solar wind features (denoted as fs), and the planetary K-index (Kp).

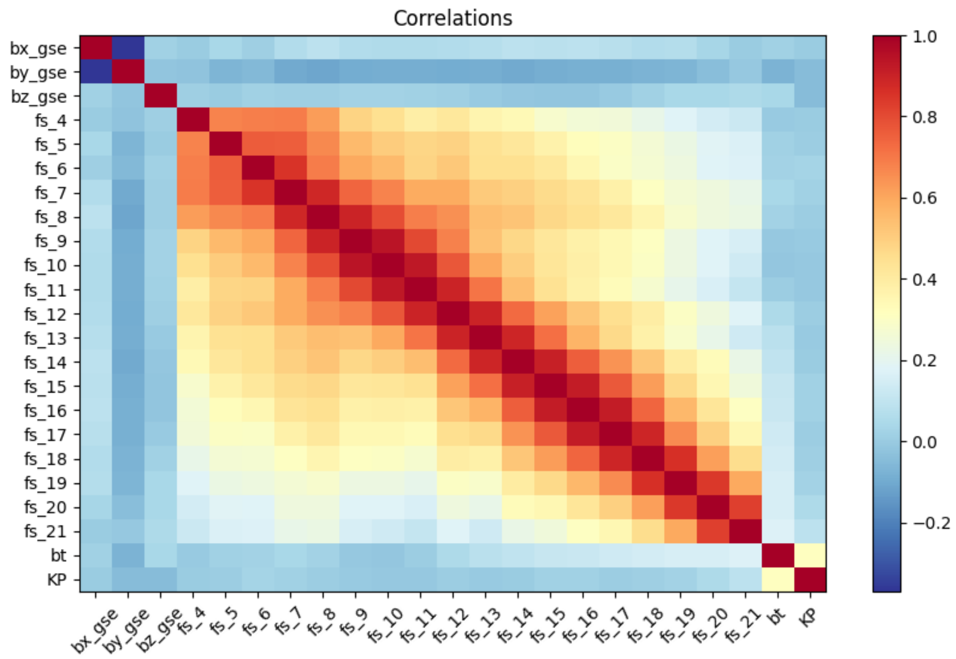

Figure 9: Correlations

8.2.1 Included Variables

1. **`bx_gse`, `by_gse`, `bz_gse`**: Components of the magnetic field in GSE coordinates.

2. **`fs_4` to `fs_21`**: Solar wind features, potentially related to the density, speed, and temperature of solar particles.

3. **bt**: Total magnetic field magnitude.

4. **KP**: Planetary K-index, which measures the magnitude of geomagnetic storms.

8.2.2 Interpretation of the Correlation Matrix

1. **Main Diagonal**:The main diagonal shows a perfect correlation (value of 1) for each variable with itself, which is expected.

2. **Significant Correlations**:

   (a) The dark red areas in the matrix indicate a high positive correlation between the corresponding variables.

   (b) A high correlation is observed among some solar wind features (`fs_6` to `fs_15`), suggesting these features are closely related to each other.

3. **Low Correlations**:

(a) The light blue areas indicate a low correlation or negative correlation between the corresponding variables.Notably, the magnetic field components (bx_gse, by_gse, bz_gse) show a low correlation with most other solar wind features.

4. **Relationship with Kp**:

   (a) The Kp index shows a notable correlation with certain solar wind features, such as fs_15 and bt. This suggests these specific solar wind parameters may be good predictors of geomagnetic activity measured by Kp.

   (b) The lower correlation with the magnetic field components (bx_gse, by_gse, bz_gse) suggests these variables have a less direct impact on the Kp index value compared to other solar wind features.

In summary, the presented correlation matrix is an essential tool for identifying key relationships between measured variables and the Kp index, facilitating the development of more accurate and efficient models for predicting geomagnetic storms.

8.3 Machine Learning Model Results

The figure 10 shows the results of a Kp classification model, evaluated using 19-fold cross-validation. The Kp values classified are 0, 50, 60, 70, and 80. Below is a detailed analysis based on the confusion matrix and the classification report.

```
Start Results Model  : 19  Fold
Confusion Matrix: [[ 89055  17759   2421    701    364]
 [ 10370  89332   8940    406    906]
 [   211   1803  85349  19559   3672]
 [     0      0    594  80622  28836]
 [     0      0      0   5306 104770]]
Report of Clasification Indicator
              precision    recall  f1-score   support

         0.0       0.89      0.81      0.85    110300
        50.0       0.82      0.81      0.82    109954
        60.0       0.88      0.77      0.82    110594
        70.0       0.76      0.73      0.74    110052
        80.0       0.76      0.95      0.84    110076

    accuracy                           0.82    550976
   macro avg       0.82      0.82      0.81    550976
weighted avg       0.82      0.82      0.81    550976

End Results Model  : 19  Fold
```

Figure 10: Classificattion Model Results

1. **Overall Accuracy**:The model's overall accuracy is 82%, indicating good performance in classifying Kp values overall.

2. **Performance by Class**:

   (a) **Class Kp=[0-40]**: With a precision of 0.89 and recall of 0.81, resulting in an f1-score of 0.85. This shows the model is very good at correctly identifying this class, although there is room for improving sensitivity.

   (b) **Class Kp=50**: With a precision of 0.82 and recall of 0.81, the f1-score is 0.82. The consistency between precision and recall suggests that the model is balanced for this class.

   (c) **Class Kp=60**: Exhibits a high precision of 0.88 but a lower recall of 0.77, resulting in an f1-score of 0.82. This indicates that while the model identifies true positives well, it misses some true instances.

   (d) **Class Kp=70**: With a precision of 0.76 and recall of 0.73, the f1-score is 0.74. This shows moderate performance and the need to improve both precision and sensitivity for this class.

   (e) **Class Kp=80**: With a precision of 0.76 and recall of 0.95, the f1-score is 0.84. The high sensitivity indicates the model is excellent at detecting this class, although precision could be improved to reduce false positives.

3. **Average Metrics**

   (a) **Macro Average**: With a precision of 0.82, recall of 0.82, and f1-score of 0.81, it reflects a good overall balance in the model's performance across all classes.

   (b) **Weighted Average**: Similar values of precision, recall, and f1-score (all at 0.82 and 0.81 respectively) suggest the model maintains its performance when considering the support of each class.

8.3.1 Considerations

1. **False Negatives and Positives**: Classes Kp=70 and Kp=80 exhibit higher numbers of false negatives and positives, indicating a need for refinement in classifying these specific categories.

2. **Class Imbalance**: Despite the seemingly balanced distribution in terms of support, performance varies significantly between classes, suggesting the need for class re-balancing approaches or additional model parameter adjustments.

3. **Future Improvements**: Enhancing sensitivity for class Kp=60 and precision for classes Kp=70 and Kp=80 could be prioritized. Methods such as collecting more data, hyper-parameter tuning, and using advanced data balancing techniques could be explored to improve these aspects.

In summary, the classification model shows good overall performance with an accuracy of 82%, but specific areas could benefit from further optimization to improve precision and recall in particular classes.

9 Conclutions

The successful integration of data from the DSCOVR and CASSIOPE missions has resulted in a robust and comprehensive dataset that significantly enhances the predictive capabilities for geomagnetic storms. By combining real-time and historical measurements of solar wind and ionospheric conditions, we have developed a highly accurate model for predicting the planetary K-index (Kp) with a lead time of 20 minutes to 1 hour. This model demonstrates an overall accuracy of 82%, marking a substantial improvement in early warning systems for geomagnetic storms.

The correlation analysis conducted between the integrated datasets and the Kp index has identified key predictors that are crucial for accurate geomagnetic activity forecasts. This has not only improved our understanding of the dynamics of space weather but also provided valuable insights into the interactions between solar wind features and Earth's magnetic field.

Our machine learning approach, utilizing a neural network model, has shown that advanced predictive analytics can effectively mitigate the impacts of geomagnetic storms on modern technologies such as GPS satellite systems and electrical power grids. The model's ability to accurately classify Kp values across different severity levels ensures that stakeholders can implement timely and appropriate measures to safeguard critical infrastructure.

Furthermore, this research highlights the importance of continuous data monitoring and advanced data processing techniques. The challenges posed by instrument sensitivity issues on the DSCOVR mission underscore the need for ongoing improvements in data collection and integration methodologies.

In conclusion, the integration of DSCOVR and CASSIOPE data has proven to be a game-changer in the field of space weather prediction. Our predictive model not only enhances early warning systems but also contributes to a deeper understanding of geomagnetic phenomena. This project sets a new standard for future research and development in geomagnetic storm forecasting, paving the way for more resilient and adaptive technological systems.

10 Disclaimer

It's important to highlight that 100% of the formulas presented in this research have been developed by the authors based in NASA SPACE CHALLENGE 2023 research. However, it's pertinent to mention that the responsibility for the correct formulation and development of these equations rests solely with the author.

11 Acknowledgments

We extend our deepest gratitude to the entire team at NASA CHALLENGE 2023 for their invaluable support and guidance throughout the research process. Their expertise and dedication have been instrumental in the successful integration of data from the DSCOVR and CASSIOPE missions. The insights and resources provided by NASA have significantly contributed to the development of our predictive model for geomagnetic storms. This project would not have been possible without their unwavering support, and we are immensely grateful for their collaboration and commitment to advancing the field of space weather prediction.