Team #: DS
Team Name:
Members: Aimen Rehman, Rithwik Kamalesh, Rochelle Daley, Tom Tran, Sadia Akther
Section: Tuesday

Introduction / Overview:

The system integrates an ensemble of models that process data from magnetograms, EUV images, X-ray time series, proton flux time series, and physical parameters from active region patches. It is engineered to assess data quality, ensure data availability, and generate accurate predictive analytics. The results are delivered via an interactive web-based interface, providing visualizations and insights into potential space weather impacts on key systems such as power grids, aviation, communications, and astronaut safety.

General description:

- Objective of the User
  - The primary objective is to predict space weather events, such as solar flares or geomagnetic storms, that can impact various sectors.
- User Characteristics
  - IT and system administrators responsible for ensuring smooth operation, data integrity, and system uptime for real-time forecasting.
- Features
  - Key features include data quality, real-time data integration, monitoring and alerts.
- Benefits
  - The system offers increased accuracy, protection of critical systems, and enables real-time monitoring.
- Why it's Important
  - Space weather can cause blackouts, disrupt GPS signals, affect aviation routes, and harm astronauts. Predicting these events allows for proactive measures to mitigate their impact.
- Target User Community

- Research Institutions: Researchers who study solar activity and its impacts on Earth.

Functional Requirements:

The data pipeline relies heavily on the cleanliness and quality of the data, ensuring that at every stage, the data format is consistent, as it's coming from different sources.

Data Prep Stage:

- Data must be cleaned continuously and in real time, as the pipeline will be pulling data from external sources at predetermined intervals.
- All data fed to the models must be compatible with all times synced to each other or grouped together to avoid inconsistencies in model output.
- Magnetograms and similar data must be packaged as JSON to ensure they can be read by the model without causing errors and inconsistencies.
- Data downloader must be connected to a data log, so any server connection problems are immediately reported, ensuring the model doesn't run without the appropriate data.

Databases:

- Requires at minimum two databases that will be storing the model outputs, and the calculated predictions.

Other:

- Bugs, crashes and any other issues must be reported to those overseeing the pipeline so appropriate action can be taken.

Interface Requirements:

- Python is required for most of the coding and data preparation.

- SQL will be used to connect models, databases, and the web interface.
- Python and Tkinter will be used to create the GUI for SWx interaction.
- CSS, JSON, JavaScript and PHP will be used to build the website interface.
- MATLAB or Tableau will be used to create the visualizations for the results.

Performance Requirements:

- Real-Time data ingestion: This system should be able to ingest and process real-time data from multiple different repositories simultaneously. This ensures that the updates happen in a timely manner, providing up to date responses and predictions, which is important for forecasting.
- Scalability: Since the forecasting data is simultaneous, the system should be able to process large data volumes. The system should support at least 10 concurrent model executions with varying data inputs without significant performance degradation.
- Response Time: The system should be able to give back prediction results to the user in a maximum of 5 seconds. This process includes fetching, staging, model interference, and displaying results. In this time span, the system should provide an accurate visualization to the user interface in a timely manner.

Design Constraints:

- Reliability: The system must run continuously to process real-time data without interruptions, making it crucial to minimize crashes and runtime errors. If one part of the system fails, backup components will ensure continuous operation without any interruptions.

- Software Dependencies: The system relies on specific software libraries, APIs, and third-party services for gathering data, running models, and creating visualizations. Changes or updates to these tools could affect the system's performance or cause compatibility issues. Careful management of these dependencies and regular updates are essential to maintain smooth operations.

- Hardware Limitations: The system requires substantial computational power and storage to handle large volumes of real-time data and run multiple models simultaneously. Limited processing power could affect the system's ability to deliver timely predictions, so ensuring sufficient hardware resources is critical.

Non-Functional Attributes:

The non-functional attributes describe how well the system performs. Non-Functional Requirements serve as the source of many functional requirements which are of immense significance and need to be specified in the right level of detail in the requirements documents. Moreover, they are essential inputs to some key architectural decisions and help ensure that some key design aspects are considered from the outset rather than during the later stages of the process when the cost and risk of change get higher.

- Data Quality: Ensuring the quality of the data is essential for accurate predictions and system performance.

- System Performance: Performance refers to the system's ability to execute its functions efficiently, using minimal resources. It is an important quality attribute as it impacts system design and hardware selection.

- Reliability and availability: Availability is the degree to which the solution is operable and accessible when required. It is a measure of time during which the system is fully operational i.e., available for use and sometimes included as a service legal agreement considering its criticality to the business. Reliability is the ability of a solution or its component to perform its required functions without failure under predefined conditions for a specified time.

- Maintainability: Maintainability refers to the ease with which a solution or its component can be fixed, enhanced, or adapted to meet business needs or changes in the environment.

- Interoperability: Interoperability is the degree to which the solution is compatible with other components. It is a measure of how effectively the system interoperates with other software systems and how easily it integrates with external hardware devices.

- Scalability: Scalability refers to the degree to which a solution can evolve to handle increased amounts of work. The increased amount of work could be in terms of the user base, transactions, data, network traffic, or other factors.

Appendices:

README File:

All technical details, including system setup instructions, software dependencies, installation steps, and troubleshooting guidelines, are provided in the accompanying README file. This file is designed to help developers and users get the system up and running smoothly and understand its architecture.

Definitions and Acronyms:

- Magnetogram: A visual representation of the sun's magnetic field strength, used to study solar activity.
- EUV (Extreme Ultraviolet) – A type of radiation observed from the sun, used for analyzing solar activity.
- Proton Flux Time Series: A measurement of the flow of protons in space, often collected by satellites, used to monitor solar activity and space weather events such as solar flares
- X Ray Time Series: A continuous record of X-ray emissions from the sun, often used to detect solar flares and other space weather events.

References:

1. NASA Solar Dynamics Observatory (SDO) - Provides data on solar activity, including magnetograms and EUV images.
   - Website: https://sdo.gsfc.nasa.gov/