**1. Project Overview**

The project plan must include the following:

**· A Project Title.**

Predictive Modeling of Solar Activity: Analyzing Daily Sunspot Data (1818–2019) Using Machine Learning and Time Series Forecasting

**· A short summary of the project topic and background.**

This project concerns prediction of solar activity before 1818 using various data science methods for prediction of daily sunspot data from 1818–2019. Key indicators of solar activity, sunspots are temporary phenomena on the Sun’s photosphere appearing as spots darker than the surrounding areas and have been systematically recorded for over two centuries. Solar activity is important for space weather of which understanding and forecasting is crucial as it can impact electronic items from satellite communications to power grids to other systems. Using machine learning techniques and time series forecasting methods, this research attempts to build models that predict the future solar activity patterns from the historical observations of sunspots.

**· A Research Question.**

Can machine learning enhance predictive accuracy beyond traditional statistical models?

**· The Project Objectives.**

* To conduct a comprehensive analysis of the historical daily sunspot data to identify trends, patterns, and anomalies that are pertinent to solar activity.
* To develop and train machine learning models using the historical sunspot data to forecast future solar activity levels.
* To assess the performance of the developed models by comparing their predictions against actual observed data, ensuring accuracy and reliability in forecasting.

**· Reference List**

Herrera, V.V., Soon, W. and Legates, D.R., 2021. Does Machine Learning reconstruct missing sunspots and forecast a new solar minimum? *Advances in Space Research*, 68(3), pp.1485–1501.

Rodríguez, J.-V., Sánchez Carrasco, V.M., Rodríguez-Rodríguez, I., Pérez Aparicio, A.J. and Vaquero, J.M., 2024. Hemispheric Sunspot Number Prediction for Solar Cycles 25 and 26 Using Spectral Analysis and Machine Learning Techniques. *Solar Physics*, 299(8), p.116. <https://doi.org/10.1007/s11207-024-02363-2>.

Velasco Herrera, V.M., Soon, W., Hoyt, D.V. and Muraközy, J., 2022. Group Sunspot Numbers: A New Reconstruction of Sunspot Activity Variations from Historical Sunspot Records Using Algorithms from Machine Learning. *Solar Physics*, 297(1), p.8. <https://doi.org/10.1007/s11207-021-01926-x>.

**2. Project Plan: Task List and/or Project Timeline**

The project is structured into key tasks, each with specific objectives and timelines, as outlined below:

| **Task Number** | **Task Description** | **Start Date** | **End Date** | **Notes** |
| --- | --- | --- | --- | --- |
| 1 | **Literature Review**: Research existing studies on solar activity prediction using machine learning and time series analysis. | Feb 10, 2025 | Feb 24, 2025 | Gather insights on methodologies and identify gaps in current research. |
| 2 | **Data Acquisition**: Obtain the daily sunspot data from the SILSO database. | Feb 10, 2025 | Feb 12, 2025 | Access data from [SILSO Data Files](https://www.sidc.be/SILSO/datafiles). |
| 3 | **Data Preprocessing**: Clean and preprocess the data to handle missing values and normalize formats. | Feb 13, 2025 | Feb 20, 2025 | Ensure data quality and readiness for analysis. |
| 4 | **Exploratory Data Analysis (EDA)**: Analyze the data to identify significant patterns and trends. | Feb 21, 2025 | Mar 6, 2025 | Utilize statistical methods and visualization tools. |
| 5 | **Model Selection and Training**: Choose appropriate machine learning algorithms and train models. | Mar 15, 2025 | Mar 31, 2025 | Experiment with various algorithms to determine the best fit for the data. |
| 6 | **Model Evaluation and Validation**: Test the models against validation datasets to assess performance. | Apr 1, 2025 | Apr 14, 2025 | Use metrics for evaluation. |
| 7 | **Results Documentation**: Compile findings, including data analysis, model performance, and insights. | Apr 22, 2025 | Apr 28, 2025 | Prepare comprehensive documentation of the project outcomes. |
| 8 | **Final Report Preparation**: Draft and finalize the project report for submission. | Apr 29, 2025 | May 5, 2025 | Ensure the report is well-structured and includes all necessary components. |

**3. Data Management Plan**

**Overview of the Dataset**

Data used is daily sunspot numbers recorded from January 1, 1818 to December 31, 2019 sourced from Sunspot Index and Long-term Solar Observations (SILSO) database of the Royal Observatory of Belgium. This data is collected primarily for monitoring solar activity over a long period.

**Data Collection**

SILSO Data Files will have direct downloading of their data. The particular file of interest is the 'Daily total sunspot number' dataset which gives the total sunspot count on a day-by-day basis within the given range.

**Metadata**

**Format**: The data is available as ASCII text and CSV.

**Size**: There are approximately 73,000 records in our dataset because that was the number of days between 1818 and 2019. The file size is usually not large, around 5MB.

**Variables**: The sunspot number for every date is included with the date.

**Document Control**

The project will be managed in the form of the dedicated GitHub repository for the codebase and the version control management. Version control and traceability will be achieved through commitment of code and documentation updates to the repository on a weekly basis.

**ReadMe File**

A comprehensive ReadMe file will be provided in the GitHub repository of the programme containing the following:

* A brief description of the project's purpose and objectives.
* What is required for setting up the development environment and dependencies.

**Security and Storage**

In order to maintain the data and code integrity and security throughout the project, the following actions will be taken:

* To avoid the data and code from being lost due to unforeseen circumstances, data and code will be backed up weekly.
* The main place we store will be at a dedicated GitHub repository making it easier for control of versions and access to other people. All files will also be stored on OneDrive provided by the University of Hertfordshire so that the storage is secure and compliant.
* Relevant staff and markers will be given access to GitHub repository for collaboration and assessment. The secure backup may be stored in OneDrive storage, and can be used for sharing the data in an internal project team.

Ethical requirements: You must address each of the following issues and state how your specific dataset meets these requirements, give evidence when possible (e.g. screenshots or references):

1. Does the data come under GDPR requirements?

The historical sunspot numbers data present in the dataset are generic, and do not contain any personal data. Consequently, GDPR is not applicable to this dataset.

2. Does the project conform to UH ethical policies?

Considering that this project does not involve human participants or personal data, the project complies with the University of Hertfordshire’s ethical guidelines. The University’s policy states that ethical approval is not required for research where the inclusion of published secondary sources is the limit.

3. Do you have permission to use the data for your proposed research project?

The SILSO database is maintained by the Royal Observatory of Belgium and data of sunspots is publicly available from there. The data is intended for public use with no limitations on use either for research purposes.

4. Are you assured that the data was collected ethical (i.e. by the original people who gathered/collected/ collated/made the data)?

Reputable scientific organisations have collected and maintained the data collected in an ethical manner that follows the standard practise in data collection and dissemination. No information is given about any ethical problems concerning the original data collection procedure.