**7PAM2002 Data Science Project Module**

**Choosing a Project Form**

**Semester B 2024/2025**

This form will be used by your supervisor to agree your project topic and dataset. Complete as much of the form as possible then submit the form into ‘Assignments’ on the Project Module Canvas site.

**Student Full Name: Ali Sufiyan**

**Student ID: 23037749**

**Email Address: gondalalisufyan31@gmail.com**

**Course: 1 year / 2 year sandwich / 2 year Advanced Research** *(delete as appropriate)*

**Semester intake to the course: A / B (***delete as appropriate)*

|  |  |
| --- | --- |
| SECTION A |  |
| **What is your proposed project title or topic?** | Predictive Modeling of Solar Activity: Analyzing Daily Sunspot Data (1818–2019) Using Machine Learning and Time Series Forecasting |
| **What is the Research Question for your project?** | Can machine learning enhance predictive accuracy beyond traditional statistical models? |
| **Dataset website address (or organisation and person the dataset is from).** | <https://www.sidc.be/SILSO/datafiles> |
| **Where was the data originally collected? (who, when, where)** | Astronomers and observatories around the world collected and coordinated their sunspot data from 1818 to 2019 (the daily data), and the Royal Observatory of Belgium serves as SILSO partner and coordinates the data for the SILSO group, including the compilations of both daily sunspot numbers (SSN) and sunspot groups (SG) which are referred to as the solar activity series. An important step occurred with systematic sunspot observations started in 1818, recorded first at observatories as in the case of the Zurich Observatory, in the 19th century under Rudolf Wolf. Both input and output datasets were standardized and extended over the course of time through the global networks such as the World Data Centre for the Sunspot Index. SILSO compiles and maintains this historical data today, and makes it available for use in scientific research and application of solar activity forecast. |
| **What type of data are you using? (e.g., image/tabular/category/continuous etc)** | The histograms displayed in this file page store the daily sunspot data, which consist of series of numerical records of sunspot counts observed through time. Each row represents a given date with columns usually including the date and columns such as that date’s corresponding daily sunspot number and uncertainty estimates. Since some datasets contain more attributes such as standard deviation, observation per day, group sunspot number, you might have some such datasets available. The data is chronologically ordered so time series analysis and trend detection could be done. This structured format allows to apply statistical methods, machine learning models, and forecasting techniques to examine long term solar activity patterns and their possible influence on space weather and climate. |
| SECTION B |  |
| Can you attend all 6 supervision sessions on-campus? If not state reason. | Actually my residential location is way ahead of campus. So It will be very difficult for me to attend weekly meetings. |
| What was the subject of your BSc degree and any other Master’s degrees you have taken? | BS in Physics |
| What are your career aims and/or the industry sector you would like to get into? | My career aims are in data science or healthcare analytics, where I can apply machine learning, statistical modeling, and data-driven insights to improve public health outcomes. Specifically, I’m interested in working in the healthcare or pharmaceutical sectors, focusing on predictive modeling for disease outbreaks, evaluating treatment effectiveness, and optimizing public health interventions. |
| A brief account of your programming/data science experience (including work or placement). | I have experience working with Python, R, and SQL, with a focus on data analysis, statistical modeling, and machine learning. I’ve worked on several projects, including predictive modeling for healthcare data, where I utilized machine learning techniques such as logistic regression and decision trees. My project on COVID-19 forecasting involves analyzing clinical trial data to assess treatment efficacy, and I’ve used tools like Pandas, NumPy, and Scikit-learn to build predictive models and visualize trends. |
| Any work experience (including non-computing related). | I’ve worked as a research assistant on healthcare-related projects, where I analyzed medical data and contributed to studies evaluating treatment outcomes. This experience helped me refine my skills in data cleaning, statistical analysis, and research methodology. |
| Do you have any hobbies, activities or interests. | I am passionate about staying updated on the latest developments in healthcare and data science And I love to play cricket and other sports activities. |