**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: Muhammad Talha Jamil**

**Student ID: 23017847**

**Email Address: talhajamilqwe@yandex.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 Analytics on the Recent California's Wildfire Risk Assessment and Damage Estimation Using Climate and Environmental Data |
| **What is the Research Question for your project?** | How do machine learning models use climate, environmental, and historical wildfire data to accurately predict wildfire risk and estimate potential damage in high-risk Californian regions? |
| **Dataset website address (or organisation and person the dataset is from).** | <https://data.ca.gov/dataset/cal-fire-damage-inspection-dins-data/resource/b8aeb030-140d-43d2-aa29-1a80862e3d62>  <https://open.fiscal.ca.gov/dept_spending_transaction.html> |
| **Where was the data originally collected? (who, when, where)** | CAL FIRE collected the data during wildfire damage assessments for the CAL FIRE Damage Inspection (DINS) Programme. Begun in 1997 under the auspices of the FIRESCOPE programme, the programme has since adapted to using Geographic Information Systems (GIS) and mobile app technologies for the collection and analysis of data. When Battalion Chief Jonathan Cox's CAL FIRE teams assess damage in wildfire-affected areas across California, it includes the 2015 Valley and Butte Fires, among other high-profile fires. Data inform wildfire risk mitigation and policy decisions. |
| **What type of data are you using? (e.g., image/tabular/category/continuous etc)** | The primary type of information for damage inspections in this CAL FIRE Damage Inspection (DINS) programme is tabular: rows and columns with variables like property address, categorical damage classifications (e.g., “destroyed,” “affected”), continuous geospatial coordinates (latitude, longitude), temporal (inspection date), etc. It has a structured format usable for interfacing with external datasets (for example, climate records or vegetation indices) to broaden further the temporal extent of the modelled wildfire risk and impact in the future. The tabular form is further enriched with geospatial data, often embedded as GIS coordinates and enriches the framework with spatial analysis and visualisation of wildfire patterns.  The programme utilises a combination of cross-validation and temporal splits for validation and testing to confirm the robustness of the model. Categorical damage label imbalances can be addressed via imbalanced cross-validation (e.g., k-fold or stratified) and generalise across different geographical regions. Dividing data by periods, or temporal validation, is important to determine estimated model performance on future wildfires as many wildland fire occurrences change with seasonality or climate-related processes. A holdout test set of unseen spatial and temporal conditions is set aside to evaluate model accuracy and ensure model reliability of policy decisions or resource allocation. Leave-one-region-out strategies also aid with the spatial validation of results by mitigating the overfitting of the localised environmental factors. Outlier detection in continuous variables, e.g., anomalous coordinates, and consistency audits for categorical labels are added as data integrity checks to strengthen the validation pipeline further. |
| SECTION B |  |
| Can you attend all 6 supervision sessions on-campus? If not state reason. | Yes, I can attend all 6 supervision sessions on-campus. I have no scheduling conflicts at the moment. |
| What was the subject of your BSc degree and any other Master’s degrees you have taken? | BS in Software Engineering |
| What are your career aims and/or the industry sector you would like to get into? | I am aiming to work as a data scientist or machine learning engineer in the environmental sector, where I can utilize my analytical skills to solve complex problems and contribute to business growth. |
| A brief account of your programming/data science experience (including work or placement). | I am aiming to work as a data scientist or machine learning engineer in the tech or finance sector, where I can utilize my analytical skills to solve complex problems and contribute to business growth. |
| Any work experience (including non-computing related). | I completed an internship where I used machine learning algorithms to analyse sales data and improve customer targeting. |
| Do you have any hobbies, activities or interests. | I’d love to read books. |