School of Arts and Sciences

Department of Computer Science

B.Sc. in Computer Science

COMP 4811: Final Year Project-1

# **Proposal Form**

|  |  |  |
| --- | --- | --- |
| Title | Comprehensive Approach For Extreme Weather Event Forecasting Based on Machine Learning and Data Science Techniques | |
| Project author: | Tariq Aziz |  |
| Supervisor: | Dr. Muhammad Fayaz |  |
| Co-supervisor (if applicable): | Dr.Azmat Hussain |  |
| Main subject Area(s): | Machine Learning, Data Science, Climate Science, Environmental Forecasting | |
| Keywords: | Extreme Weather, Machine Learning, Forecasting, Climate Change, Avalanches, Landslides, Droughts | |
| Project type: | Research-based Project | |
| Methodologies: | 1. Data Collection  2. Data Cleaning and Preprocessing 3. Machine Learning Techniques (Supervised Learning: Classification)  4. Statistical Analysis and Data Visualization 5. Model Evaluation and Validation  6. Web app development | |
| Short project description: | This project aims to develop a machine-learning model to predict extreme weather events (landslides, avalanches, and droughts) by combining historical climate data and environmental factors. The objective is to create a forecasting tool to predict extreme weather events, allowing for better preparation and mitigation strategies. The project will leverage various machine learning techniques to analyze datasets from global climate agencies and forecast potential extreme weather scenarios. The project seeks to address the growing need for predictive analytics in the face of increasing climate variability due to global climate change. | |
| Project Aim and Objective(s): | * Develop Predictive Models: Design and implement a machine learning pipeline to predict avalanches, landslides, and droughts based on real-time meteorological data with greater accuracy. * Integrate Data from Multiple Sources: Integrate data from multiple sources, like past data and weather forecast data. * Web-Based Dashboard: Create a user-friendly dashboard that displays a predictions section, level of risk, selection of geographical region, geospatial maps, and interactive charts, as well as historical trends. * Safety Recommendations: The system would provide safety tips and recommendations based on the predictions. * Scalable and Reliable Deployment: Deploy the system on scalable platforms for continuous accessibility and performance. | |
| Equipment and critical resources required: | 1. High-performance computing resources (for model training) 2. Climate datasets (e.g., NOAA, Kaggle) 3. Machine Learning libraries (e.g., Scikit-learn, XGBoost, etc.) 4. Visualization tools (e.g., Matplotlib, Plotly) | |
| Recommended pre-requisites / Knowledge required and Supporting 3rd Year Study units: | 1. Machine Learning 2. Data Science 3. Programming in Python 4. Climate Science (basic understanding) 5. Data Visualization techniques | |
| Foreseeable Ethical issues and how these will be tackled (if applicable): | No major ethical concerns. However, proper citation of climate data sources and adherence to open-source licensing will be ensured. | |
| Copyright note: (specify copyrights) | The model and code produced in this project will be copyrighted under open-source licenses where applicable, but any datasets from external sources will retain their original copyright as specified by the data providers. | |
| Expected outcomes: | 1. Machine learning models capable of forecasting extreme weather events (avalanches, landslides, and droughts). 2. A web app and dashboard to showcase predictions and visualization results. 3. A comprehensive report detailing the development and evaluation of the model. | |
| Expected deliverables:  **Note**: The Project Proposal and report must be signed by the author, supervisor(s), and department chair. | 1. Project Proposal: (original file + pdf) soft copy + hard copy 2. Project Report: (original file + pdf) soft copy + hard copy 3. Presentation (original + PDF) 4. GitHub repository as a downloaded zip file 5. Supervisor(s) review 6. External expert review 7. Plagiarism report 8. Product (installation file, link to deployed website) | |
| Estimated Budget in USD: | $250 | |
| Language support: | English | |
| GitHub repository[[1]](#footnote-2) link: | https://github.com/tariqaxix/Final\_Year\_Project | |
| Programming language(s): | Python | |
| Framework (if applicable): | Scikit-learn, XGBoost, Django | |
| External libraries: | NumPy, Pandas, Matplotlib, Plotly | |

**Note**: In case you implement a commercial project for the company, you must provide the company’s consent to publish the project report by UCA. As you can see, the project proposal includes lots of sensitive information, so it is important to get the client’s consent before the project starts.

**Note**: A proposal that lacks signatures is rejected as it lacks legal power.

Submission date: 23/ 10/ 2024

**Important**: This page is used for department purposes only and must be filled by faculty only.

# Department committee approval:

**Approve / Reject**

Dr. Ayman Aljarbouh (signature) :

Chair of the Computer Science department

Date : \_\_\_\_\_\_\_\_\_\_\_ December 2022

**NOTE**: This document is evolving. It will be subjected to revisions with a view to help the current and future students learn and enjoy more from their FYP experience. If you think of relevant points, welcome to let me know and I'll add in your observations. A completed project proposal can simply be cut and pasted into your report later.

Table of Contents

[Comprehensive Approach For Extreme Weather Event Forecasting Based on Machine Learning and Data Science Techniques 1](#_Toc197120913)

[Project Description 8](#_Toc197120915)

[Literature Survey 8](#_Toc197120916)

[Proposed Solution 10](#_Toc197120917)

[Development Approach 11](#_Toc197120918)

[Proposed Work 11](#_Toc197120919)

[Business Benefits 12](#_Toc197120920)

[Contributions 12](#_Toc197120921)

[Similar applications comparison 12](#_Toc197120922)

[Technical specification of the project 13](#_Toc197120923)

[Functional requirements 14](#_Toc197120924)

[Nonfunctional requirements 14](#_Toc197120925)

[Project Budget Estimation 16](#_Toc197120926)

[Project plan and schedule 17](#_Toc197120927)

[Milestones 17](#_Toc197120928)

[Gantt Chart 18](#_Toc197120929)

[Risk management plan 19](#_Toc197120930)

[Risk Assessment Table 19](#_Toc197120931)

[Risk Solutions 20](#_Toc197120932)

[References: 21](#_Toc197120948)

# **Project Description**

Climate change has been attributed to a rise in extreme weather conditions like avalanches, landslides, and droughts, which have devastating impacts on human life, infrastructure, and ecosystems. In return, these events have become not only more frequent but also more severe. Hence, the ability to predict such occurrences becomes more important than ever as the unpredictability of weather patterns not only hampers government planning of disaster management but also puts several industries at high risk: agriculture, energy, and insurance.

This project, entitled Extreme Weather Event Forecasting- Utilizing Machine Learning and Data Science, would go a long way in using historical climate data and applying machine learning techniques to come up with strong predictive models. The ultimate objective is to accurately forecast landslides, avalanches, and droughts, providing a dependable tool for governments, businesses, and communities to prepare and respond to such events.

Several scenarios justify the importance of this project. For instance, imagine a scenario where avalanche-prone regions in Central Asia can be given accurate early warnings, allowing for timely evacuation and infrastructure protection. In agriculture, a drought prediction tool could enable farmers to take preventive actions to reduce crop loss, thereby securing food supplies and stabilizing economies. Another critical scenario involves energy management—power grids could optimize operations if extreme weather events were forecasted well in advance.

By providing precise forecasts, this project will contribute to minimizing the damage from extreme weather events, thereby supporting sustainable development and disaster resilience efforts globally.

# **Literature Survey**

Predicting extreme weather events has long been a critical area of research. However, due to climate change, the need for a robust and efficient disaster forecasting system has increased. Camps‐Valls et al. (2025) explain that the impact of climate change is observable in the growing intensity and frequency of weather events, which are detrimental to the environment and humanity. These events include long periods of drought, severe floods, destructive typhoons, and avalanches. Methods of disaster forecasting based on statistics, including numerical weather prediction (NWP), are obsolete. They have profound inadequacies because of the intricate, ever-changing, and non-linear interrelations that exist among the components of the climate. Recently, advances in technology such as deep learning methods have shown vast potential in the prediction as well as analysis of extreme events (Camps‐Valls et al., 2025). Through the application of various algorithms, it is possible to identify and interpret multiple relationships and trends within big datasets. Unfortunately, datasets containing meteorological information are often heterogeneous, noisy, and limited, which impacts the performance of the algorithms.

A well-optimized forecasting model learns from data from various sources to capture inherent climate characteristics. Data from multiple sources, such as topographical, atmospheric, and meteorological observations, form typical data used to train a model. Heterogeneous data are consistently integrated into studies: Examples include soil moisture, hydrological variables, and atmospheric variables, in the context of drought forecasting (Nandgude et al., 2023). Such data can be obtained from IOT devices, satellites, and weather stations (Nandgude et al., 2023). Despite having a vast amount of data, a significant challenge comes in integrating data effectively. Current studies indicate that a vast amount of data is not being used to enhance accuracy in forecasting disasters (R. Chen et al., 2020). ML algorithms perform flawlessly when learned using such rich datasets. For example, studies have discovered that ML model accuracy grows by incorporating remote sensing data in landslide forecasting (Akosah et al., 2024).

Machine learning, more specifically, supervised machine learning, has widespread applications in forecasting extreme phenomena. For instance, river water level or drought index can be considered a regression problem, whereas forecasting or classification of the category would be a classification example. Mosavi et al. (2018) showcased the potential of ML algorithms like support vector machine (SVM), Decision Trees, Random Forest, and Artificial Neural Network in flood forecasting more accurately. Some other studies have shown the superior performance of ML models over classical statistical models. For instance, Nandgude et al. (2023b) used a number of ML models to forecast drought in Algeria, in which SVM performed better than other models with a coefficient of determination of 0.95. Similarly, tree-based ensemble models along with logistic regression have also produced encouraging results in flood as well as landslide predictions. Supervised ML models get an advantage over simplistic statistical models on account of being able to identify subtle relationships between predictor and outcome variables. With a limitation of needing plenty of data to train upon, these models are not always possible when rare disasters are to be forecasted. The datasets are generally unbalanced (few “extreme” instances vs a lot of normal instances). Techniques like cross-validation, feature selection, and hyperparameter tuning prove to be helpful in such a situation to ensure generalization.

Recent applications of machine learning models in disaster forecasting indicate an increased intersection of AI technology and extreme weather forecasting. Multiple trends are promoting the use of AI to forecast extreme phenomena. One of those trends includes developing “Trustworthy AI” for forecasting weather and climate phenomena. Beyond simple forecasting, Generative models such as GANs (Generative Adversarial Networks) and VAEs (Variational Autoencoders) are employed to enhance data. To demonstrate this, Camps-Valls et al. (2025b) employed VAE to enhance the prediction of heatwaves by generating various variations of patterns of heatwaves. Additionally, interdisciplinary collaboration illustrates the involvement of climate scientists, hydrologists, engineers, computer experts, as well as policymakers. All these examples show the rise of AI in extreme event forecasting.

Central Asia is subject to various disasters like drought, earthquakes, avalanches, and landslides. These countries have been using traditional methods of forecasting for a long period of time. Over recent years, this region has seen a boost in the implementation of AI in weather and climate forecasting. Sadrtdinova et al. (2024b) employed machine learning classifiers (logistic regression, random forests, gradient-boosted trees) and a deep learning model to forecast drought. Their predictive models achieved an impressive 97-99% accuracy at a 6-month lead time. Xu et al. (2024) also came up with a machine learning model that was able to forecast 33% of flash drought. With this, international agencies like the World Bank are incorporating ML into drought monitoring platforms. These developments are a giant leap towards using AI techniques for disaster forecasting.

The magnificent Tien Shan, as well as the Pamir mountain range, are found in Central Asia. These mountain areas are prone to snow avalanches. Scientists are attempting to find these high-risk areas and forecast avalanches well in advance. As an example, Rahmati et al. (2019) employed ML models to identify the likelihood of avalanche activity in two mountain ranges (the Iranian Alborz and a section of the European Alps). Similarly, Wen et al. (2022) were able to identify avalanche hazard areas by using ML techniques in the high-altitude Tibetan Plateau region. In a similar work, Bian et al. (2022) utilized an ensemble classifier to enhance avalanche hazard mapping accuracy. Based on earlier work, Choubin et al. (2020) employed four algorithms: generalized additive model (GAM), multivariate adaptive regression splines (MARS), boosted regression trees, and support vector machines (SVM) to determine avalanche predictive features. Among these, their models were highly accurate to a level of more than 88%. It reflects the potential of ML models to identify these terrain, snowfall, and climate interactions that contribute to avalanches effectively.

The geographic nature of Central Asian, particularly Tajikistan and Kyrgyzstan, makes them prone to landslides. Landslide susceptibility assessment in these areas by traditional methods fails because of the vast extent of geographic areas involved, along with remote sites. Machine learning models are filling this gap by using, as well as learning from, various data (e.g., rain, geology, topography, earthquakes) (Rosi et al., 2023). In a pioneering work, Rosi et al. (2023) produced the first detailed landslide susceptibility map covering the entire region. They employed a dataset of 13,000 historic landslide reports to train a random forest model that could forecast areas of landslide susceptibility. Several other authors have also contributed a lot to this field of work. For instance, X. Chen et al. (2024) developed a “Tien Shan–Kunlun Mountains Landslide Susceptibility Model (TKLSM)” to estimate the risk of landslides in data-scarce areas. This model estimated an annual casualty rate of 3–4 in Kyrgyzstan and Tajikistan, much lower than in remote regions (X. Chen et al., 2024b). Overall, Random Forest and other ML models have shown great results in landslide susceptibility mapping​ in the region. More work needs to be done on the quality of data and data integration from other sources for a robust and reliable landslide forecasting model in Central Asia.

Despite the increase in ML use for disaster predictions, several research gaps remain. For instance, most countries still rely on experts for avalanche prediction. This indicates the hesitation in applying ML and data-driven methods for hazard forecasting. Secondly, most of the existing ML-integrated systems hardly evaluate the performance of different algorithms. Researchers have noted that “only a few studies” compare different ML models for drought predictions (Oyounalsoud et al., 2024). Another major challenge is the class imbalance in hazard datasets. Avalanches and landslides rarely occur compared to non-hazardous instances. This skewness highly impacts the learning of the model as it gets accustomed to the non-hazardous instances. Furthermore, the lack of models trained on Central Asian data implies that models developed elsewhere are used in the Central Asian context. Acharya et al. (2023) state that Central Asian mountains have “so far remained inadequately studied” in avalanche forecasting. The models used in these contexts are trained in Europe or North America, so they don’t capture the unique geographical and climate conditions. Central Asia is a “data-scarce region,” thus making it even harder to train a model locally for drought or any other disaster prediction (Pyarali et al., 2022).

In conclusion, we have seen great advancement in the use of Machine Learning for droughts, avalanches, and landslides forecasting in Central Asia. Different models, such as Support Vector Machine (SVM) and Random Forest, have shown promising results in extreme weather forecasting. Government and organizations are switching to ML for better hazard predictions, early warnings, and risk assessments. This contributes to the body of knowledge and finds practical implications in risk reduction strategies. As climate change increases the frequency and intensity of natural hazards, ML-integrated systems play a pivotal role in safeguarding communities and infrastructure.

# **Proposed Solution**

This project aims to address some of these challenges by developing a comprehensive machine-learning model capable of predicting a wide range of extreme weather events (e.g., floods, droughts, heatwaves) using datasets from NOAA, Kaggle, and other global climate agencies. The use of various models will allow us to experiment and choose the one with the best performance. Moreover, the project will implement a user-friendly interface for the visualization of forecasts. This solution will provide critical early warnings, improving disaster preparedness for regions vulnerable to extreme weather events.

In summary, while the field of machine learning in weather forecasting has made significant strides, the development of multi-event forecasting models remains a critical area of research. By building on the existing body of work, this project will contribute to the development of a robust and comprehensive weather forecasting tool that can assist in the global effort to mitigate the impacts of climate change.

# **Development Approach**

The development approach selected for this project is Agile, given the need for flexibility and iterative improvements. Agile allows frequent updates to the model, dataset, and performance metrics, ensuring that technical and business aspects of the project are continuously refined. This approach suits machine learning-based projects as it encourages experimentation and adjustments based on the results from model training and validation phases.

Agile is appropriate for this project due to the following reasons:

* Iterative Nature: The model needs to be fine-tuned and evaluated continuously, making it easier to implement improvements.
* Flexibility: Agile allows us to respond quickly to new data or unforeseen issues, such as data quality or model underperformance.
* Collaboration: Frequent reviews can improve the model’s accuracy, resulting in better predictions.
* Continuous Feedback: Agile encourages frequent testing and feedback cycles, which are essential in machine learning projects to refine the model and results.

The system will be evaluated on:

* Model Accuracy: Using evaluation metrics like precision, recall, F1-score, and confusion matrices to assess performance.
* User Interaction: Evaluating the user interface’s usability through focus groups and testing.
* Deployment and Performance: Assessing the reliability and response time of the deployed system under real-world conditions.

# **Proposed Work**

The thesis will focus on developing an extreme weather event forecasting system using machine learning. Key work items include:

* Data Collection: Sourcing climate data from NOAA, Kaggle, and other global agencies.
* Model Development: Designing and implementing machine learning algorithms for prediction.
* Evaluation: Conducting rigorous tests on model performance.
* User Interface: Developing a web-based interface for visualization.
* Deployment: Hosting the model and UI on cloud platforms.

# **Business Benefits**

* Improved Disaster Preparedness: The focus of the project is to offer early warnings and hazard evaluations to assist local communities and agencies.
* Informed Decision-Making: The information and forecast provided by the system would help policymakers and the government make appropriate decisions in the event of any emergency. Typical policies may include resource allocation, evacuation planning, and emergency response strategies.
* Cost Reduction: Timely predictions can reduce the costs associated with disaster relief efforts. By better preparation, the system can prevent or minimize the impact of disasters, thus reducing the associated recovery costs.
* Increased Public Awareness: The user-friendly web application would allow users to analyze the risks and take preventive measures. This can significantly reduce the loss in various sectors such as agriculture, energy, and insurance.
* Scalability and Adaptability: The system can be expanded to handle more disasters, making it an ideal choice in various geographical regions. Additionally, the system can be extended to process real-time data from satellites and other sources.

## **Contributions**

This project can contribute extensively to the field of extreme event forecasting, risk assessment, and disaster management. It will assist students and researchers in using machine learning techniques for disaster predictions. Additionally, the project involves the common mass and makes it convenient for them to access the disaster data. Moreover, this project would be available on my GitHub and would serve as a prototype for researchers and students for further research. There are a lot of disaster forecasting apps, but very few of them utilize machine learning techniques. Furthermore, most of these apps target a specific disaster, thus lacking the multi-forecasting ability. Our app would not only allow the prediction of multiple disasters but would also leave room for the integration of other disaster data in the future. Apart from being a robust app, it will be a reliable and cost-effective solution for governments, local agencies, and emergency responder teams.

# **Similar applications comparison**

**Note**: Provide a comparison table of similar applications/projects by comparing the features of the applications/projects. Indicate all features to be implemented within your project by providing the complete list. The depth and quality of your comparison are going to affect your grade for the Midterm exam. Do not include features that are very common or present in all applications. Describe features that require some

clarification/explanation. Provide links to the applications’ websites for validation.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Features** | **NASA Earth Exchange** | **UN-SPIDER Risk Tool** | **GFDRR GeoNode** | **My App** |
| Predicts multiple disasters | Drought, Flood | Flood, Landslide | Earthquake, Flood | Drought, Avalanche, Landslide |
| Uses Machine Learning | Deep Learning | Rule-based | GIS-driven | RF, LR, Stacking |
| Disaster-specific tips for end users |  |  |  | Contextual safety tips |
| Real-time weather integration |  |  |  | OpenWeatherMap |
| Interactive web dashboard |  | Limited UI | Map UI | Full dashboard |
| Risk prediction using local weather input |  | From satellite |  | Form-based manual inputs |
| Free and open-source | Yes | Yes | Yes | Yes |

Sources:

* <https://www.nasa.gov/nasa-earth-exchange-nex/>
* <https://www.un-spider.org/risks-and-disasters>
* <https://www.gfdrr.org/en/onlinetools>

# **Technical specification of the project**

* Frontend: HTML5, CSS3, Bootstrap 5, JavaScript
* Backend: Python 3.11, Django 4.x
* ML Frameworks: Pre-trained models using Scikit-learn, XGBoost, LightGBM, CatBoost
* Libraries: Pandas, Numpy, Joblib
* Data Storage: SQLite3 (Django ORM)
* Visualization: Matplotlib, Seaborn, and Plotly Express for interactive graphs
* External APIs: OpenWeatherMap API for current weather data
* Hosting Environment: A dedicated server to host our web app

## **Functional requirements**

FR1: City Weather Fetch

Description: When we enter a city, the system should get real-time weather data (temperature, humidity, wind speed, etc.) for it.

Acceptance Criteria: When a user enters a city's name, it fetches the weather data within 2 seconds.

FR2: Disaster Type Selection

Description: The user should have the option to select from three disasters.

Acceptance Criteria: The drop-down menu shows all three disasters.

FR3: Dynamic Form Rendering

Description: The system should dynamically display the environmental features upon selecting a disaster.

Acceptance Criteria: The correct form appears with appropriate fields.

FR4: Risk Prediction

Description: The system should use the pretrained models to make predictions.

Acceptance Criteria: The system returns the prediction with a safety tip within 3 seconds.

FR5: Data Logging

Description: The system should save the prediction data, such as city, coordinates, disaster type, and risk level, in the database.

Acceptance Criteria: Correct entries appear in the admin panel.

FR6: Visualization Dashboards

Description: The dashboard should display multiple data visualizations

Acceptance Criteria: All the maps and charts render without any error.

FR7: Responsive Interface

Description: The system should support both desktop and mobile versions.

Acceptance Criteria: Elements resizing on devices with a width < 768px.

## **Nonfunctional requirements**

NFR1: Performance

Description: The application should be responsive and return the output within an acceptable latency.

Acceptance Criteria: API response time < 3s, ML prediction time < 2s.

NFR2: Scalability

Description: The system should be able to support more disasters in the future.

Acceptance Criteria: New models can be added without much change.

NFR3: Usability

Description: The system should be user-friendly.

Acceptance Criteria: Testing shows more than 90% user satisfaction.

NFR4: Reliability

Description: The system should be up 99% of the time.

Acceptance Criteria: Downtime results in alerts and cached content.

NFR5: Cross-Browser Compatibility

Description: The system should work identically on all types of browsers.

Acceptance Criteria: Identical layout and performance on Chrome, Firefox, and Safari.

NFR6: Accessibility

Description: Mild visual impairments should be able to use the system.

Acceptance Criteria: Font sizes >14px, color contrast ratio >4.5:1, keyboard navigability ensured.

NFR10: Responsiveness to Invalid Input

Description: Invalid or incomplete forms should not be processed further.

Acceptance Criteria: Warnings and errors on missing fields.

# **Project Budget Estimation**

**Note**: As per UCA policy senior students who are taking *Final Year Project 1 and 2* courses can use **$300** budget for different expenses directly related to the implementation of the project. Provide project budget estimation including all expected expenses during the project implementation. Use the following categories for the below table: Software, Hardware, Deployment, Training, Reserve, Online Plagiarism check.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Category* | *Description* | | *Quantity* | | *Unit price* | *Total cost* | *% Of Budget* |
| Software, Hardware, and Tools | Required software licenses, hardware resources, and tools for development | | 1 | | $ 50.0 | $ 50.0 | 20% |
| Cloud Services | Google Cloud/AWS for model training, storage, and deployment | 1 | | $ 50.0 | | $ 50.0 | 20% | |
| Hosting and Domain | Hosting and domain services for deploying the project’s web interface | | 1 | | $ 100.0 | $ 100.0 | 40% |
| Miscellaneous Expenses | Backup storage, unexpected expenses, additional tools/software | | 1 | | $ 50.0 | $50.0 | 20% |
| Total Budget Cost Estimate: $250 | | | | | | | 100% | | |

# **Project plan and schedule**

**Note**: Provide the project schedule using the Gantt Chart and comment where applicable. Identify milestones. A milestone is a concrete event that one can use to demonstrate progress. Milestones should be clear, concrete, demonstrable achievements (“SMART”).

## **Milestones**

|  |  |  |  |
| --- | --- | --- | --- |
| Milestones | Work to be done | Outcome | Deadline  dd/mm/yyyy |
| Milestone 1 | Data collection and preprocessing | Clean, prepared datasets | 09/12/2024 |
| Milestone 2 | Model design and initial implementation | Working prototype of predictive models | 08/02/2025 |
| Milestone 3 | Model evaluation and optimization | Optimized model with acceptable accuracy | 05/03/2025 |
| Milestone 4 | Development of the web app and dashboard | The functional user interface for predictions | 21/03/2025 |
| Milestone 5 | Final testing and deployment | Fully tested and deployed project | 16/04/2025 |
| Milestone 6 | Final Report, documentation, and Presentation | Complete report and documentation | 30/04/2025 |

## **Gantt Chart**

**Note**: In case your Gannt chart is very big to fit into single page you can submit separate excel file or insert screenshot of your chart in landscape orientation. Pay attention how tasks are grouped under the project phases. Provide link to online publicly available Gantt Chart in case your image is too big to fit into the page.

A screenshot of a project

Description automatically generated

# **Risk management plan:**

To manage risks effectively, I will apply the following four stages of risk management:

1. Risk Identification: I will identify potential risks, such as data availability, model underperformance, or cloud resource limitations.
2. Risk Assessment: I will assess the likelihood of these risks occurring and their potential impact on my project.
3. Risk Mitigation: I will develop contingency plans, such as using multiple data sources or adjusting model parameters if the model’s accuracy is low.
4. Risk Monitoring: Throughout the project, I will continuously monitor these risks and adjust my strategies as necessary.

## **Risk Assessment Table**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Risks | | Likelihood | Consequence | Overall Risk | Risk Level |
| (1-3) | (1-5) | (1-15) |
| 1 | High computational costs | 1 | 1 | 3 | Insignificant |
| 2 | Delays in data processing | 1 | 2 | 5 | Minor |
| 3 | Model Underperformance | 2 | 3 | 8 | Significant |
| 4 | Data Availability or Quality | 2 | 4 | 10 | Major |
| 5 | Lack of Correlation Between Events | 3 | 5 | 14 | Severe |

## **Risk Solutions**

## High Computational Costs (Insignificant Risk):

## Utilize cloud computing resources for scalable processing.

## Optimize algorithms to reduce resource consumption.

## Delays in Data Processing (Minor Risk):

## Streamline data processing pipelines for efficiency.

## Adopt efficient data storage solutions.

## Schedule regular check-ins to ensure adherence to timelines.

## Model Underperformance (Significant Risk):

## Implement cross-validation techniques to assess model effectiveness.

## Experiment with various algorithms and hyperparameters to enhance performance.

## Data Availability or Quality (Major Risk):

## Conduct thorough preliminary data quality assessments.

* Finding data from sources like NOAA and Kaggle.

## Lack of Correlation Between Events (Severe Risk):

## Perform comprehensive exploratory data analysis (EDA) to understand relationships in the data.

## Guide effective feature selection based on EDA findings to improve model accuracy.

# **References**

1. Camps-Valls, G., Fernández-Torres, M., Cohrs, K., Höhl, A., Castelletti, A., Pacal, A., Robin, C., Martinuzzi, F., Papoutsis, I., Prapas, I., Pérez-Aracil, J., Weigel, K., Gonzalez-Calabuig, M., Reichstein, M., Rabel, M., Giuliani, M., Mahecha, M. D., Popescu, O., Pellicer-Valero, O. J., . . . Williams, T. (2025). Artificial intelligence for modeling and understanding extreme weather and climate events. *Nature Communications*, *16*(1). <https://doi.org/10.1038/s41467-025-56573-8>
2. Nandgude, N., Singh, T. P., Nandgude, S., & Tiwari, M. (2023). Drought prediction: A comprehensive review of different drought prediction models and adopted technologies. *Sustainability*, *15*(15), 11684. <https://doi.org/10.3390/su151511684>
3. Chen, R., Zhang, W., & Wang, X. (2020). Machine Learning in Tropical Cyclone Forecast Modeling: A review. *Atmosphere*, *11*(7), 676. <https://doi.org/10.3390/atmos11070676>
4. Akosah, S., Gratchev, I., Kim, D., & Ohn, S. (2024). Application of Artificial intelligence and remote sensing for landslide detection and prediction: Systematic review. *Remote Sensing*, *16*(16), 2947. <https://doi.org/10.3390/rs16162947>
5. Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood Prediction Using Machine Learning Models: Literature review. *Water*, *10*(11), 1536. <https://doi.org/10.3390/w10111536>
6. Sadrtdinova, R., Perez, G. a. C., & Solomatine, D. P. (2024b). Improved drought forecasting in Kazakhstan using machine and deep learning: a non-contiguous drought analysis approach. *Hydrology Research*, *55*(2), 237–261. <https://doi.org/10.2166/nh.2024.154>
7. Xu, L., Zhang, X., Wu, T., Yu, H., Du, W., Zhang, C., & Chen, N. (2024). Global prediction of flash drought using machine Learning. *Geophysical Research Letters*, *51*(21). <https://doi.org/10.1029/2024gl111134>
8. Rahmati, O., Ghorbanzadeh, O., Teimurian, T., Mohammadi, F., Tiefenbacher, J. P., Falah, F., Pirasteh, S., Ngo, P. T., & Bui, D. T. (2019). Spatial modeling of snow avalanche using machine learning models and Geo-Environmental factors: comparison of effectiveness in two mountain regions. *Remote Sensing*, *11*(24), 2995. <https://doi.org/10.3390/rs11242995>
9. Wen, H., Wu, X., Liao, X., Wang, D., Huang, K., & Wünnemann, B. (2022). Application of machine learning methods for snow avalanche susceptibility mapping in the Parlung Tsangpo catchment, southeastern Qinghai-Tibet Plateau. *Cold Regions Science and Technology*, *198*, 103535. <https://doi.org/10.1016/j.coldregions.2022.103535>
10. Bian, R., Huang, K., Liao, X., Ling, S., Wen, H., & Wu, X. (2022). Snow avalanche susceptibility assessment based on ensemble machine learning model in the central Shaluli Mountain. *Frontiers in Earth Science*, *10*. <https://doi.org/10.3389/feart.2022.880711>
11. Choubin, B., Borji, M., Hosseini, F. S., Mosavi, A., & Dineva, A. A. (2020). Mass wasting susceptibility assessment of snow avalanches using machine learning models. *Scientific Reports*, *10*(1). <https://doi.org/10.1038/s41598-020-75476-w>
12. Rosi, A., Frodella, W., Nocentini, N., Caleca, F., Havenith, H. B., Strom, A., Saidov, M., Bimurzaev, G. A., & Tofani, V. (2023). Comprehensive landslide susceptibility map of Central Asia. *Natural Hazards and Earth System Sciences*, *23*(6), 2229–2250. <https://doi.org/10.5194/nhess-23-2229-2023>
13. Chen, X., Wang, Y., Wang, X., Li, Y., Qi, J., & Lin, Q. (2024). Risk Assessment of Landslide Casualty under Incomplete Information——Tienshan and Kunlun Mountainous Regions of Central Asia. *International Journal of Disaster Risk Reduction*, 105057. <https://doi.org/10.1016/j.ijdrr.2024.105057>
14. Oyounalsoud, M. S., Yilmaz, A. G., Abdallah, M., & Abdeljaber, A. (2024). Drought prediction using artificial intelligence models based on climate data and soil moisture. *Scientific Reports*, *14*(1). <https://doi.org/10.1038/s41598-024-70406-6>
15. Pyarali, K., Peng, J., Disse, M., & Tuo, Y. (2022). Development and application of high resolution SPEI drought dataset for Central Asia. *Scientific Data*, *9*(1). <https://doi.org/10.1038/s41597-022-01279-5>

1. provide at least “read” access to **nurlanotek** in case your repository is private [↑](#footnote-ref-2)