ML Project Sprint Plan

≔	Course	CSCI 4622
📅	Date	@October 23, 2024
✳️	Status	In progress
≔	Type	

Sprint 1: Project Setup and Data Collection

Timeline: October 24 - November 3

Objectives:

- Set up the development environment.
- Collect and preprocess datasets.
- Conduct exploratory data analysis (EDA).

Tasks:

1. Set up the project repository and libraries for machine learning.
2. Collect the EM-DAT dataset and any additional financial data.
3. Clean and preprocess the dataset (handle missing values, normalization).
4. Perform EDA to understand patterns in the dataset (disaster type, region, etc.).
5. Identify and prepare relevant features for each of the models (time series, classification, regression).

Deliverables:

- Clean, preprocessed datasets.
- EDA report and identified features for the models.

Sprint 2: Disaster Risk Prediction Model Development

Timeline: November 3 - November 13

Objectives:

- Build time series and classification models to predict disaster occurrences and types.

Tasks:

1. **Time Series Model**:

   - Implement **ARIMA** to forecast disaster occurrences based on past data.

   - Implement **LSTM** for long-term disaster trend prediction, focusing on disaster timing and locations.

   - Evaluate the performance (Mean Absolute Error, RMSE) of both models.

2. **Classification Models**:

   - Develop a **Random Forest** classifier to predict disaster types based on historical features.

   - Implement **Gradient Boosting Machine (GBM)** to improve classification accuracy and minimize error.

   - Perform hyperparameter tuning (cross-validation, grid search) to improve model accuracy.

3. Validate both models using past disaster data (split dataset into training and test sets).

Deliverables:

- Disaster occurrence prediction models (ARIMA, LSTM).

- Disaster type classification models (Random Forest, GBM).

- Model evaluation report (precision, recall, accuracy).

Sprint 3: Resource Allocation Optimization

Timeline: November 13 - November 23

Objectives:

- Implement regression and decision-making models to optimize resource allocation.

Tasks:

1. **Regression Models**:

   - Implement **Multiple Linear Regression** to estimate resources required based on disaster type, region, and magnitude.

   - Implement **Ridge/Lasso Regression** for improved accuracy and regularization (prevents overfitting when features are correlated).

2. **Decision Tree Models**:

   - Develop **CART (Classification and Regression Trees)** to determine the most important disaster features influencing resource needs.

3. **Reinforcement Learning**:

   - Implement **Q-Learning** for resource allocation optimization based on past responses, adjusting allocation strategies for future disasters.

4. Test models using historical data (cross-validation) to predict resource needs and allocation efficiency.

Deliverables:

- Resource allocation prediction models (Regression, CART, Q-Learning).

- Evaluation report on the performance and accuracy of resource prediction.

Sprint 4: Financial Impact Estimation

Timeline: November 23 - December 3

Objectives:

- Implement models to predict the financial impact of disasters and detect outliers.

Tasks:

1. **Support Vector Machines (SVM)**:

   - Implement **SVM** to identify financial outliers (e.g., disasters that result in unexpectedly high or low financial damage).

2. **Regression Models**:

- Build **Elastic Net Regression** for predicting financial impact based on disaster type, region, and population affected.

- Implement **Polynomial Regression** to model non-linear relationships between disaster magnitude and financial damages.

3. **Neural Networks**:

- Implement **Feedforward Neural Networks (FNN)** to capture complex relationships in financial data and improve the accuracy of damage predictions.

4. Validate models with real-world financial impact data (train-test split).

Deliverables:

- Financial impact prediction models (SVM, Elastic Net, Polynomial Regression, FNN).

- Model performance metrics (R-squared, Mean Absolute Error).

Sprint 5: Model Integration and Testing

Timeline: December 3 - December 13

Objectives:

- Integrate all the models into a unified system.

- Test the complete system using disaster scenarios and historical data.

Tasks:

1. Integrate the disaster risk prediction models, resource allocation models, and financial impact estimation models.

2. Test the integrated system with different disaster scenarios (e.g., floods, earthquakes) using historical data to simulate real-world responses.

3. Evaluate overall system performance (accuracy, speed, resource utilization).

4. Identify areas for fine-tuning and perform necessary adjustments to improve system robustness.

Deliverables:

- Fully integrated disaster prediction and response system.

- Final evaluation report on model accuracy, performance, and efficiency.

Sprint 6: Final Review and Documentation

Timeline: December 13 - December 16

Objectives:

- Final review of all models and documentation.

- Complete project presentation and submit all deliverables.

Tasks:

1. Conduct final testing of the integrated system to ensure all models work as intended.

2. Write detailed documentation, including model descriptions, implementation details, and final performance metrics.

3. Prepare a final presentation for the project, highlighting the use of machine learning models for disaster prediction, resource allocation, and financial impact estimation.

4. Submit final project deliverables (code, data, reports).

Deliverables:

- Final project documentation and presentation.

- Submission of the entire codebase, data, and reports.