ARTIFICIAL INTELLIGENCE

Project Title: *Earthquake Prediction Model*

Phase 1 : Problem Definition and Design Thinking

Problem Definition :

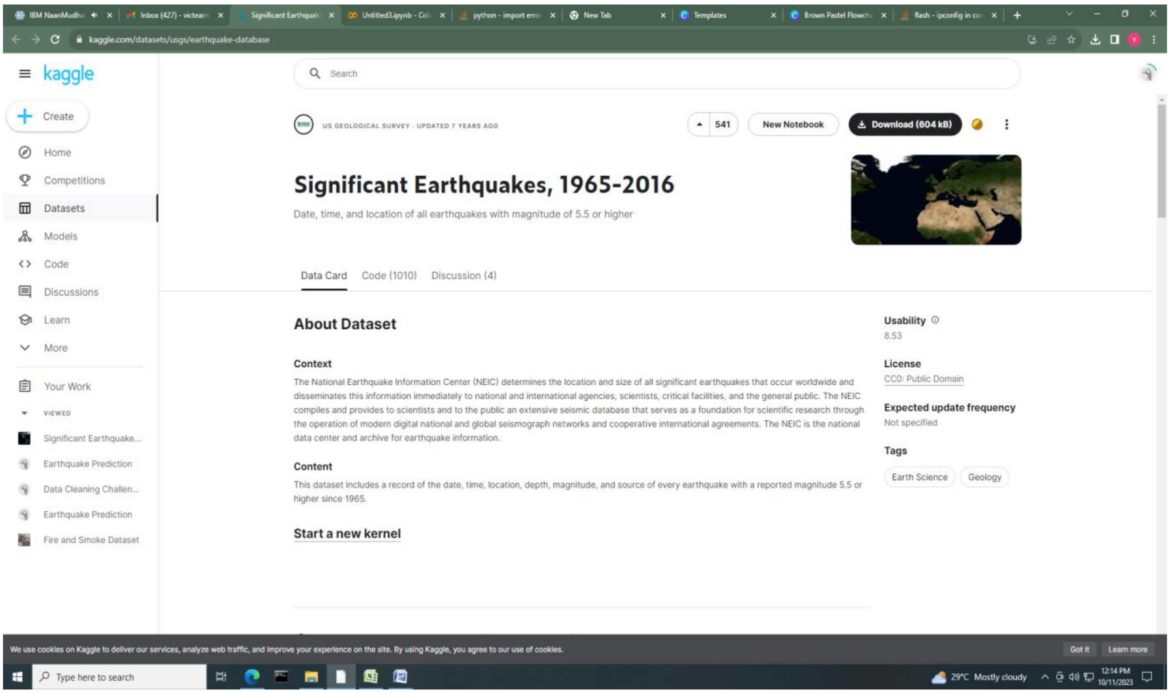
The problem at hand is to develop an earthquake prediction model using a kaggle dataset. The primary objective is to explore and understand the key features of earthquake data, visualize the data on a world map for a global overview, split the data for training and testing, and ultimately construct a neural network model that can predict earthquake magnitudes based on the provided features.

DESIGN THINKING

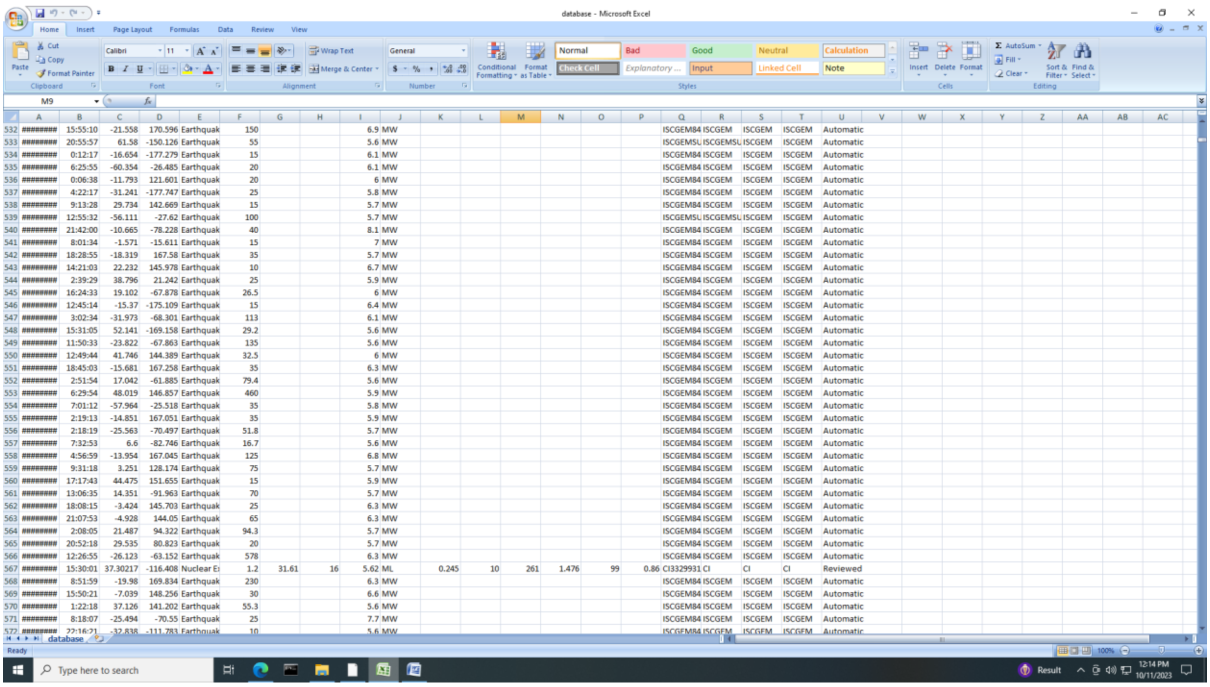
Data Source

The first step in solving this problem is selecting a suitable kaggle dataset that contains earthquake data. This dataset should include essential features such as date, time, latitude, longitude, depth, and magnitude. The choice of the dataset is crucial as it forms the foundation of our model.

Dataset Source :



Sample Data Snapshot :



FEATURE EXPLORATION

Once the dataset is acquired, it's essential to dive into feature exploration. This phase involves:

1. **Data Inspection:**

Carefully examining the dataset to understand its structure, data types, and any missing values.

2. **Statistical Analysis:**

Calculating summary statistics, including mean, median, standard deviation, and quartiles for each feature. This will help us identify outliers and understand the data's distribution.

3. **Correlation Analysis:**

Investigating the correlations between features, especially between earthquake magnitude and other variables. Identifying highly correlated features can be beneficial for model development.

VISUALIZATION

Visualization plays a crucial role in gaining insights from the data. In this phase:

1. **World Map Visualization:**

Creating a world map visualization to display the geographical distribution of earthquakes. This can help identify earthquake-prone regions and patterns.

2. Time Series Plots:

Visualizing the earthquake data over time to detect any temporal trends or seasonality.

DATA SPLITTING

To evaluate our model effectively, we need to split the dataset into two subsets:

1. Training Set:

This set will be used to train our neural network model. It should contain a significant portion of the data, ensuring that the model learns from a diverse range of examples.

2. Test Set:

The test set is crucial for evaluating the model's performance. It should be separate from the training data and used to assess how well the model generalizes to unseen earthquake data.

MODEL DEVELOPMENT

In this phase, we focus on building the earthquake prediction model using a neural network. Key steps include:

1. Data Pre processing:

Preparing the data for model input, which may involve normalization, scaling, or encoding categorical variables.

2. Neural Network Architecture:

Designing the architecture of the neural network. This includes defining the number of layers, neurons, activation functions, and loss functions.

3. Model Training:

Training the neural network on the training set using appropriate optimization techniques, such as stochastic gradient descent (SGD) or Adam.

TRAINING AND EVALUATION

The final phase involves training the model and evaluating its performance:

1. Model Training:

Fit the neural network to the training data and monitor its convergence. Adjust hyper parameters as needed to optimize performance.

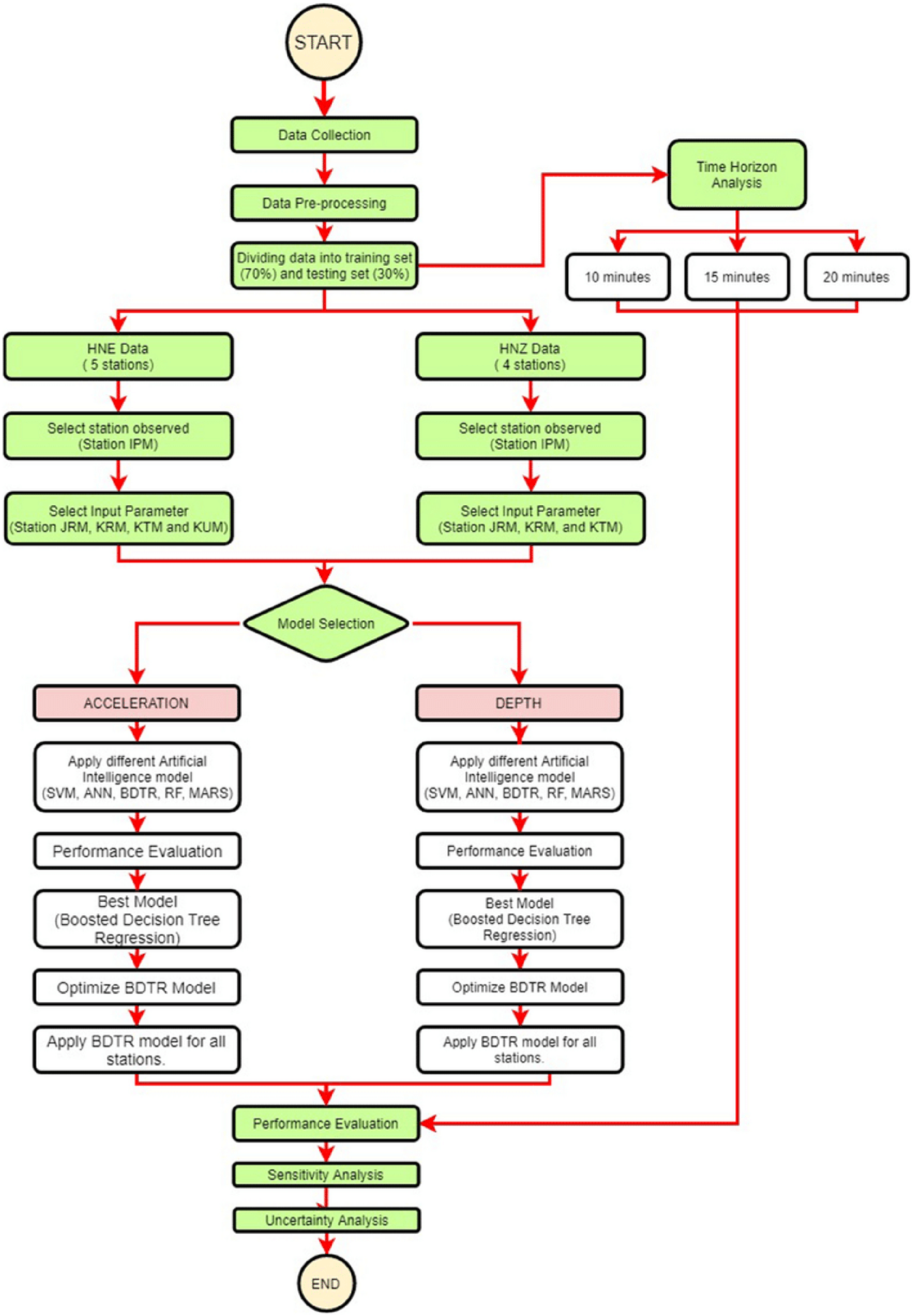
2. Model Evaluation:

Assess the model's performance on the test set using appropriate evaluation metrics, such as mean squared error (MSE) or root mean squared error (RMSE).

3. Fine-Tuning:

If the model's performance is not satisfactory, consider fine-tuning the architecture or exploring advanced techniques like hyper parameter tuning or different neural network architectures.

Flow Chart:


START
Data Collection
Data Pre-processing
Dividing data into training set (70%) and testing set (30%)
Time Horizon Analysis
10 minutes
15 minutes
20 minutes
HNE Data (5 stations)
HNZ Data (4 stations)
Select station observed (Station IPM)
Select station observed (Station IPM)
Select Input Parameter (Station JRM, KRM, KTM and KUM)
Select Input Parameter (Station JRM, KRM, and KTM)
Model Selection
ACCELERATION
DEPTH
Apply different Artificial Intelligence model (SVM, ANN, BDTR, RF, MARS)
Apply different Artificial Intelligence model (SVM, ANN, BDTR, RF, MARS)
Performance Evaluation
Performance Evaluation
Best Model (Boosted Decision Tree Regression)
Best Model (Boosted Decision Tree Regression)
Optimize BDTR Model
Optimize BDTR Model
Apply BDTR model for all stations.
Apply BDTR model for all stations.
Performance Evaluation
Sensitivity Analysis
Uncertainty Analysis
END

CONCLUSION

In summary, the project aims to develop an earthquake prediction model by following a systematic approach, from data acquisition to model evaluation. The key is to combine data analysis, visualization, and machine learning techniques to create a robust model capable of predicting earthquake magnitudes. Regular monitoring and iterative improvements will be essential throughout the project's lifecycle to achieve the best possible results.