# Natural Disaster Analysis

In order to analyse and to see the effectiveness of the environmental calamities as well as the victims of disasters, the effective data has been collected from EM-DAT (The International Disaster Database), was taken from its website www.emdat.be. The data collected was fromyears 1900 to 2021. Then, the information obtained was measured through different data mining processes to track useful information and insights from the large level data scale and the data mining processes including regression, classification, outlier detection, clustering, sequence analysis, association rules, social network analysis, time series analysis, sentiment analysis etc.

**Research Question**: Future prediction of Natural Disaster

- For this research, the data features are transformed and scaled for better analysis, the data is then used for time series forecasting using the ARIMA model. The dataset may contain some missing values. This treatment can be done in four ways – remove the rowscontaining missing values, remove the columns containing missing values, imputing the missing values in a numerical column with its mean or median and imputing the missing values in a categorical column with its mode. Outliers are treated in two ways – remove the rowscontaining outliers, replace the outliers in a column with its median value.
- The next step is to transform and scale the features in the dataset. Feature pre-processing is a crucial step in development of a Machine learning or ML model. It is the only way to gain a better score and also a crucial point how you represent your data and feed it to a target model. There are various techniques like Min Max Scaling, Robust Scaling, Log Transformation, Max Abs Scaling, Label Encoding and many more is used.
- Finally, the data is used for forecasting and time series analysis. In these natural disasters analysis, we use the data from 1900 – 2021 and try to forecast the natural disasters in the future i.e., from the year 2022, using the ARIMA model. ARIMA, stands for Autoregressive Integrated Moving Average Model with three order parameters:*(p, d, q).*

**Database work:** We have created a database table in Azure, *Natural_Disaster_Event* is a table name, and we inserted 1000+ records and 7 important columns for Natural Disaster Analysis. And merged this data into master table *Master_Disaster_Events* for further visualization and analysis of global Natural Disaste*r.*

**Challenges:** After fitting SARIMAX() model, the code prints out its respective AIC score. Because some parameter combinations may lead to numerical misspecifications, we had to explicitly disabled warning messages in order to avoid an overload of warning messages.