Datasets

Primarily, I should define what is the datasets. The datasets are collection of data that are used to train and test the Artificial intelligent algorithms and forecast. The datasets can be into structure or unstructured data which depend on the being solved. The environment and climate data can be divided on four categories. First the remote sensing data which is the one that came from satellite like NASA’s MODIS. Second, weather station data which are temperature, rainfall, wind and humidity. Third, IoT sensors networks which are air quality sensors, soil moisture sensors, river flow meters. Four, open datasets which are meteorological organization, NOAA, and Copernicus. There are two categories of algorithms is used to process the datasets. First, machine learning which take satellite data and sensor readings. For example, support vector machines (SVM) which is good for classification rainfall prediction. In addition, random forests which are aggregate models that split data into decision rules, which widely used for prediction events. Second, deep learning which also take big complex datasets (satellite imagery and IoT streams). Deep learnings algorithms contain three common algorithms used in prediction, Convolutional Neural Networks (CNNs) ,and Recurrent Neural Networks (RNNs) especially LSTM (Long Short-Term Memory), and Hybrid CNN-LSTM Models. For the first, CNN is Specialized in spatial data (satellite images, weather radar) and Great for precipitation mapping, cyclone tracking, air pollution maps. Second, RNN and LSTM is designed for time-series forecasting, and handling temporal dependencies (e.g., multi-day rainfall, drought onset). Third, CNN-LSTM models which combine spatial and temporal learning for tasks like dust storm prediction, flood forecasting.(O’Gorman & Dwyer, 2018)