Project ID: 31210

Project Name: Advanced Cyclone Intensity Prediction System using Deep learning

Python version: Python 3.11.11

Python Libraries used:

1. Numpy
2. Pandas
3. Matplotlib
4. Seaborn
5. Sci-kit learn
6. Tensorflow

Dataset used: RSMC-IMD Pune Best track Dataset (1983-2024)

Difficulties faced during project:

1. Finding a relevant dataset was one difficulty we faced. We had 3 options, using api call to make our own dataset, RSMC-IMD Pune Dataset and China Meteorological Survey’s Northwest Pacific Ocean Dataset. We opted for IMD Pune dataset which covered all the cyclones happened from 1983-2024 in the Arabian sea and Bay of Bengal
2. The dataset we found was high quality, but it was divided into different sheets for each year, and also had inconsistent values in some columns, like there would be a sentence describing the condition after each cyclone data, and some columns had 70% missing values. Cleaning up took a longer than expected.
3. Training all models (LSTM, GRU, RNN) with long sequences of cyclone data was computationally expensive. The complexity of processing sequential data, combined with the need to analyze multiple window sizes (ranging from 5 to 30-time steps), led to long training times even when using a GPU.
4. Handling large sequences required significant RAM and VRAM, leading to out-of-memory (OOM) errors, especially when training on a local machine. The dataset contained thousands of data points, and expanding window sizes further increased memory requirements.

To overcome the hardware limitations, we shifted the model training to Google Colab, which provided superior computational power with access to free GPUs and TPUs. This significantly reduced training time, prevented memory crashes, and allowed us to train larger models without hardware constraints.