SOLAR ENERGY PREDICTION USING ANN

Ankit Choudhary 1301CS53
Vishal Chauhan 1301CS46

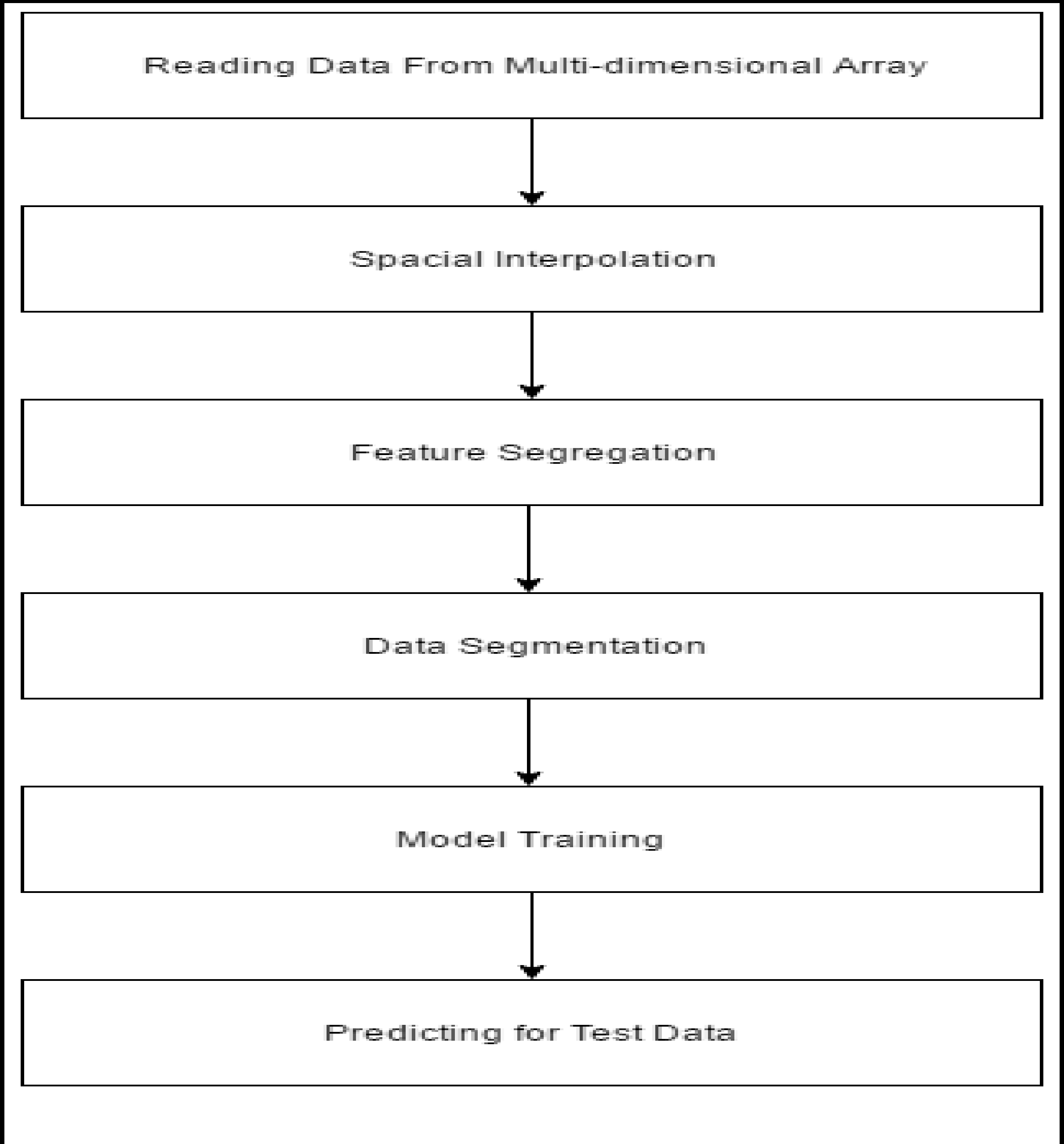
PROBLEM STATEMENT

- Solar energy production largely depends on the solar energy received at the solar installations.
- There is an increasing need for precise forecasting of amount of solar energy received at a particular location.
- Important for planning of availability of alternative energy resources.
- Held as a contest organised by AMS on kaggle.

INPUT DATASET

- Solar energy production largely depends on the solar energy received at the solar installations.
- The goal is to establish relationship between 15 given parameters and solar energy received using FFNN model.
- The training and test data provided by kaggle.com as a part of contest.
- 98 sites are marked for which predictions is to be done.
- The meteorological data comes from the GEFS and is used to predict solar energy at the marked sites.

OUR APPROACH



READING DATA

- 15 weather variables are identified as those affecting the solar energy received.
- Forecast value for each variable is given as a 5-d array in netCDF format.
- The five dimensions are date, ensemble model, forecast hour, latitude, and longitude.
- That is we have the forecast value by each of the ensemble members, at 5 different time steps on specified latitudes and longitudes.

FEATURE SEGREGATION

- We reduce the 5d matrix into 4d matrix by taking an average over all different weather prediction models (ensembles).
- To obtain a relation between these weather variables and target variable, we adopt two ways to get features
- First, using value at each forecast hour as a feature, giving 15x5 values per day
- Second, taking the average of the forecast hour values, giving 15x1 values per day

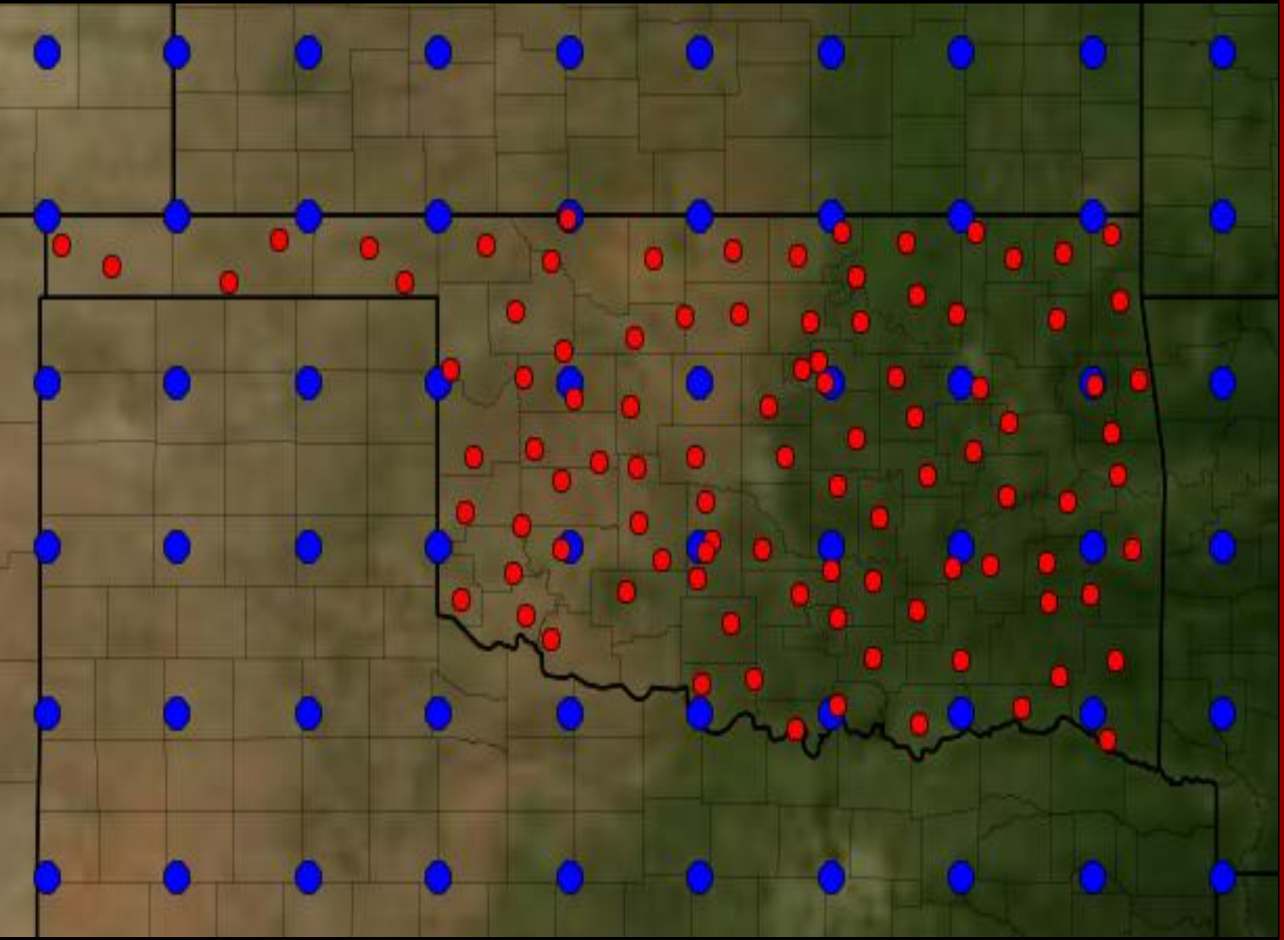


Fig1 - mesonet vs gefs points (source: kaggle)

SPATIAL INTERPOLATION

- The forecast values for the weather variables is given for the points marked in blue (fig1) .
- The Solar energy prediction is to be made for the points marked in red (fig1).
- We need to get forecast values at the intended locations.
- We use spatial interpolation using cubic splines.

DATA SEGMENTATION

- Solar energy received follows a seasonal pattern.
- For example, more solar energy is received in winters as compared to summers.
- So the data separated out season wise, may reduce the error value by training different networks for different seasons.
- For this purpose, we identify two types of segmentation:
  - Single segment with all 365 days (usual case).
  - Four segments with each having similar 91 days of the year.

NEURAL NETWORK

- We construct two different neural networks having 75 and 15 features respectively.
- We choose to use two hidden layers having 15 and 11 nodes respectively (This combination gave the best result) .
- We used PCA(principal component analysis) to reduce features from 15 to 7 and 75 to 10.
- Data segmentation was applied on 15 feature neural network, so we had a single segment run as well as four segments(season wise) run on the same.

NEURAL NETWORK

- The ‘tanh’ transfer function was used at both the hidden layers.
- We used ‘sgd’ as the training algorithm.
- Mean absolute error (MAE) was considered as the cost function.
- Maximum epochs were set to 2000.

RESULTS

Model	Train MAE	Test MAE
FFNN(15 features, single data segment)	6324693	6963159
FFNN(75 features, single data segment)	6085464	6654795
FFNN(15 features, four data segments)	4130633	4601898

Clearly, performance improved with data segmentation

RESULTS

Mean Absolute error VS Epochs

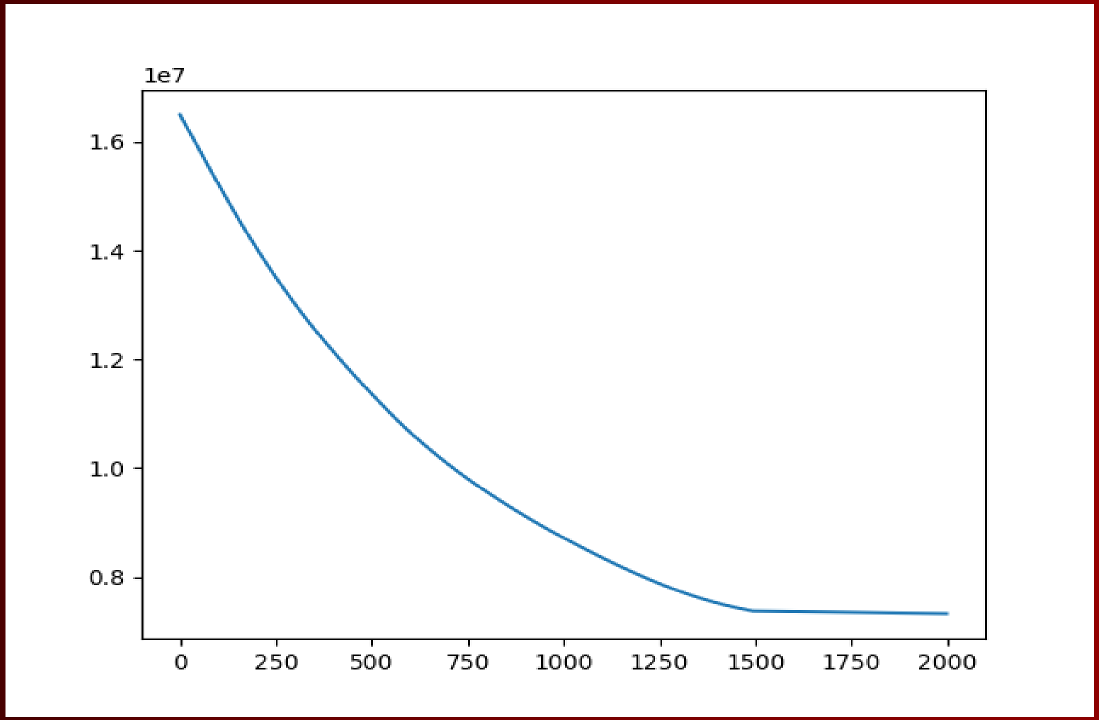


Fig2. - 15 features(single segment)

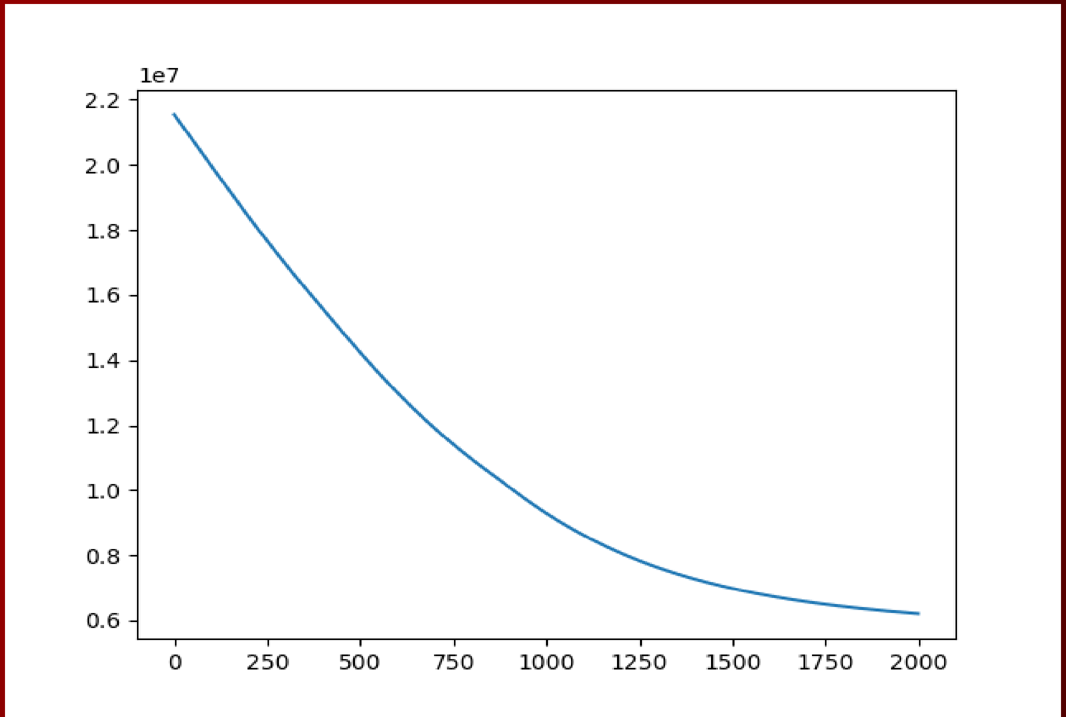


Fig3. - 75 features(single segment)

RESULTS

Mean Absolute error VS Epochs

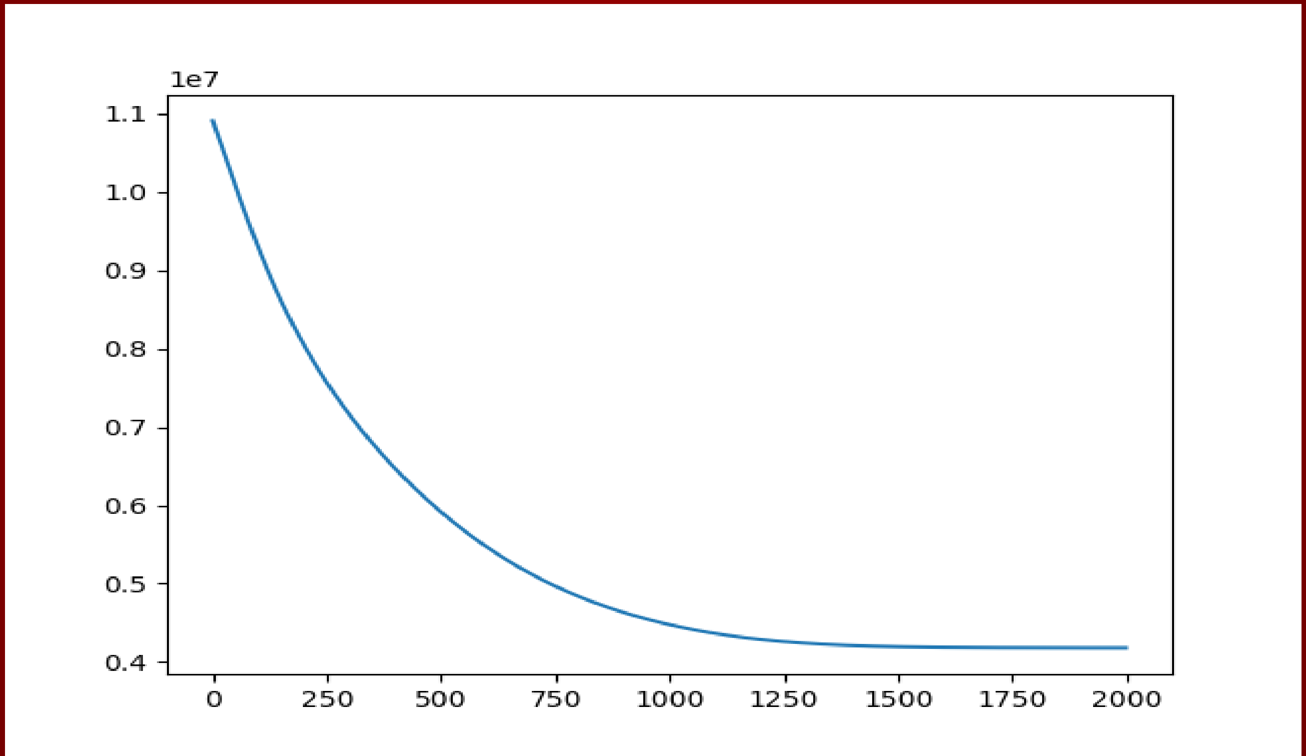


Fig4. - 15 features(4 segments)

FUTURE WORK

- There is a possibility of reducing the training time by exploring different training algorithms
- Data can be further segmented to monthly data, which may reduce the mean absolute error further.
- Sites specific data segmentation can be done to get better results.