Project Report

Mehul Kumar Nirala
16CS30046

Aryan Singh
17MT30024

Arza Akhil
17MT30004

Satyendra Nagar
19NA10026

Extreme rainfall event forecasting using neural network
Machine Learning and Earth Science

Supervisor: Adway Mitra

Table of Contents

Lists of Figures

Introduction

In forecasting rainfall, satellite imageries, ground observation stations and weather balloons were used mostly. There is also radar imaging technology, but it is not yet an extensive application due to the large image data.

Sequentially, a temporal interval of 12 hours is set to do rainfall forecast. Many scholars have focused on improving the monitoring and forecasting of weather and developing the technology to improve the ability to forecast heavy rainfall events.

In recent years, the use of artificial intelligence algorithm for rainfall forecasting has attracted considerable attention. The mechanism of rainfall forecasting is a nonlinear system in terms of mathematics.

Due to the rampant cases of cutting down of trees, the protective effects of trees against flooding are slowly being removed. Therefore, the prediction of extreme rainfall events becomes even more important. We can do two types of rainfall event forecasting :-

- Predicting rainfall over a few weeks or months in advance which is called Long Term predictions.
- Predicting rainfall a few days in advance for a given area which is called Short Term predictions.

First, we'll do some Exploratory Data Analysis on the given data for observing patterns and trends of rainfall and use feature selection techniques if required. Then, we'll make a model which will make the use of neural networks to predict the amount of rainfall for the specific states in India using the data set mentioned in the previous section. On the basis of the predicted amount of rainfall, we can classify the rainfall event as an extreme or normal.

With this data with more variations of average rainfall, it is very difficult for a statistical model to predict the required data point.Here we implement neural networks to predict the extreme rainfall, the neural net is used to create multiple features that helps in predicting the data points with more seasonal variations.

Literature Survey

At present, large numbers of models are involved in finding out possible combinations of predictors for long range forecasting; only few models with best skill are selected.The National Center for Medium Range Weather Forecasting (NCMRWF), an Indian government agency, provides daily weather and rain forecasts based on physics-based models, but the models are not capable of predicting heavy rainfall accurately Khaladkar RM, 2007.

Numerical weather prediction (NWP) models which are based on dynamical weather equations are used to provide short-range forecasts based on the present weather conditionsJC, 2004. Abundance of literature is available for predicting rainfall over Indian catchments Bhowmik RS, 2010, Sahai AK, 2000 Researchers have attempted to link the extreme rains with eccentric atmospheric behaviorHart RE, 2001; it has been observed that extreme rainfall events have a great dependence on anomalous weathers.

The fingerprinting technique, developed by **Root**, is a recently developed statistical method that uses clustering technique to detect the atmospheric variables and areas undergoing significant changes during extreme events.

Nayak et al., 2004 used the fuzzy theory to predict the outflow of the river in the Mandala watershed in India. Teschl et al., 2007 improved the forecast of weather radar with feed-forward neural networks. Some models use finite-difference methods for all three spatial dimensions, while other global models and a few regional models use spectral methods for the horizontal dimensions and finite-difference methods in the vertical dimensions JC, 2004.

Artificial Neural Networks are a proved and efficient method to model complex input–output relationships R. Aliev, 2000. The problem of estimating ground rainfall using radar measurements aloft has already been researched with neural networks. **Xiao** applied a Backpropagation Neural Network (BPNN) for rainfall estimation from radar data.

H. Liu, 2001 developed a Radial Basis Function Neural Network (RBFNN) to estimate ground rainfall using the vertical profile of the radar reflectivity as input vector. **Li** showed that the radar reflectivity from 1 to 4 km height above the rain gauge is the best input vector to a RBFNN for estimating the ground rainfall compared to several other choices. G. Xu, 2005 also used this vertical profile as input vector to their RBFNN. **Teschl** as well defined four superposed radar reflectivity measurements as input vector and added the highest level where precipitation was detected as an additional parameter.

Dataset

We will be using subdivision wise Rainfall and its departure from 1901 to 2015. These data sets are available on government website Open Govt Data (OGD) Platform India.

Indian Government has undertaken many research studies to analyze the impact of global warming and climate change on rainfall pattern in India. The analyses were made using observed rainfall data from more than 3000 rain-gauge stations spread over the country for 115 years (1901-2015). The major inferences from these studies based on the 115 years of rainfall data are as follows: The analysis of 115 years of monsoon rainfall data suggests that there is no long term change or trend in the monsoon rainfall averaged over the country. Even though, there are no changes in the all-India rainfall, there are significant changes in annual rainfall in some meteorological sub-divisions. Rainfall over Kerala, East Madhya Pradesh, Jharkhand, Arunachal Pradesh and Nagaland, Manipur, Mizoram and Tripura (NMMT) show decreasing trends. However, rainfall over coastal Karnataka, Maharashtra and Jammu and Kashmir show an increasing trend.

3.1 Data characteristics

There is a general tendency of increasing frequency of extreme rainfall (heavy rainfall events) over India, especially over the central parts of India during the southwest (June- September) monsoon season. There is no evidence of global warming on the observed changes in annual or seasonal rainfall over India. However, there is growing evidence suggesting that increasing frequency of extreme rainfall is due to global warming. The climate change assessment made by the Intergovernmental Panel on Climate Change (IPCC) suggest that in future, frequency of extreme rainfall may increase over India due to increase in global warming. However, there are NO other long term changes/trends in rainfall over India which can be attributed to global warming. The Indian Monsoon is found to be a stable system. With this data with more variations of average rainfall, it is very difficult for a statistical model to predict the required data point.

From the last plot, it is clear that if the amount of rainfall is high in the months of July, August, September then the amount of rainfall will be high annually. It is also observed that if amount of rainfall in good in the months of October, November, December then the rainfall is going to be good in the overall year.

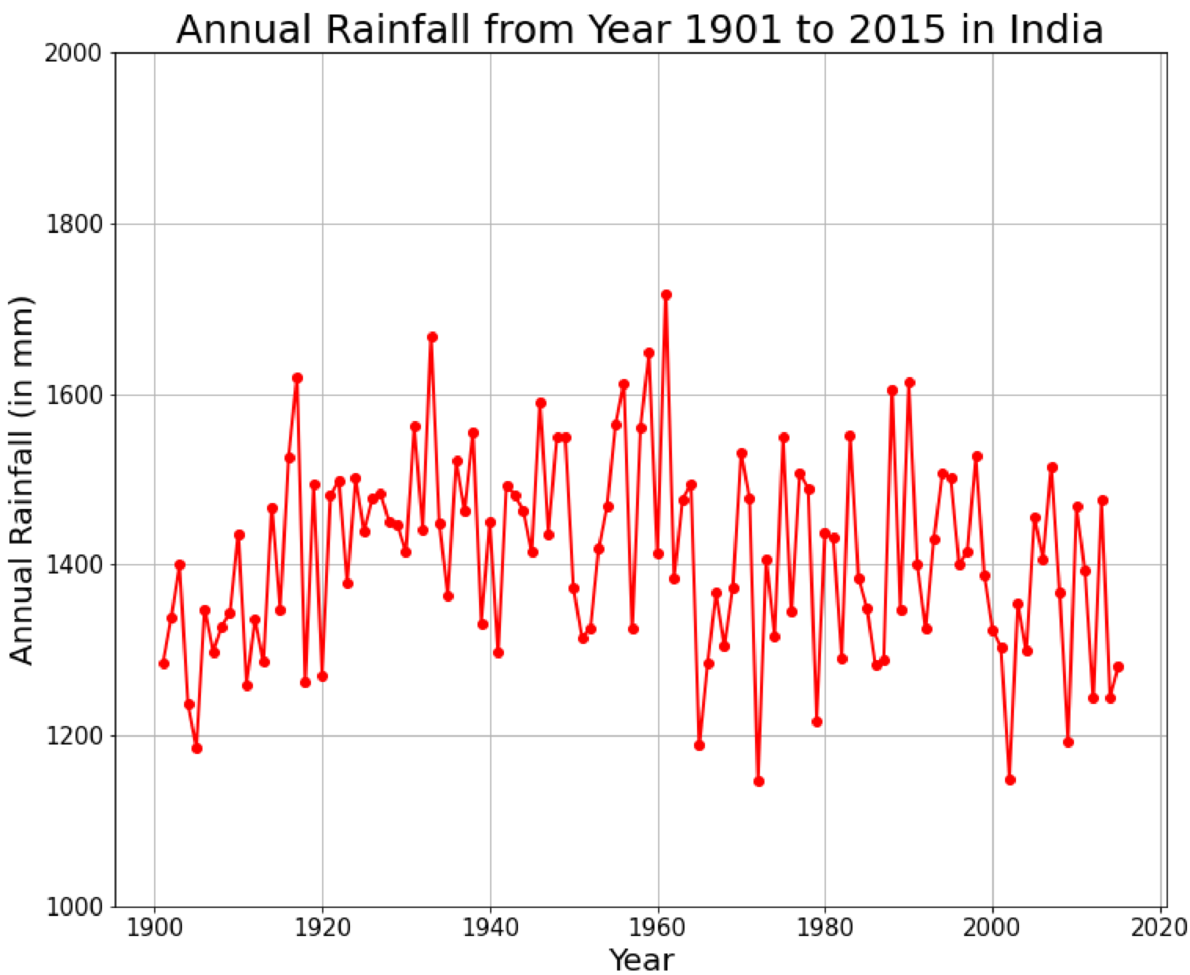


Figure 3-1: Annual Rainfall from 1901-2015 in India

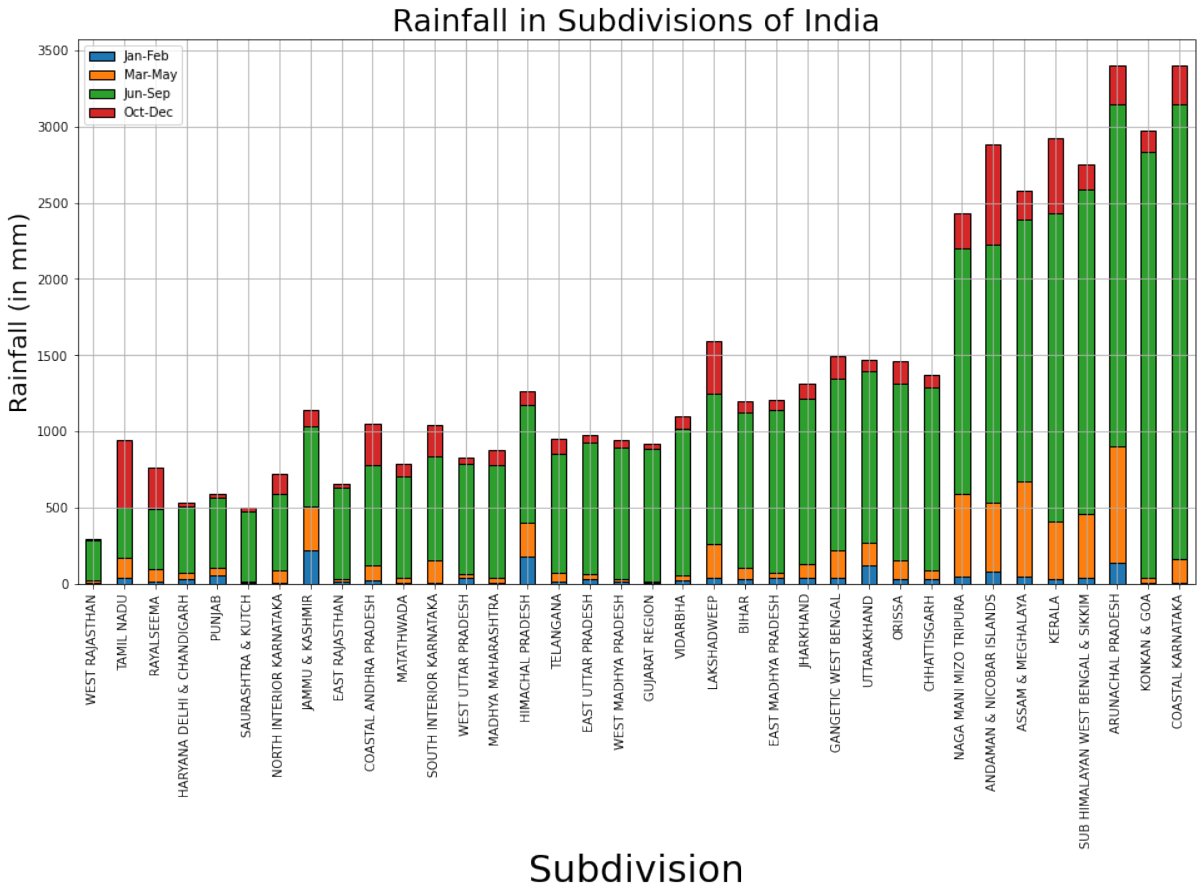


Figure 3-2: Subdivision wise Annual Rainfall from 1901-2015 in India for a period of months.

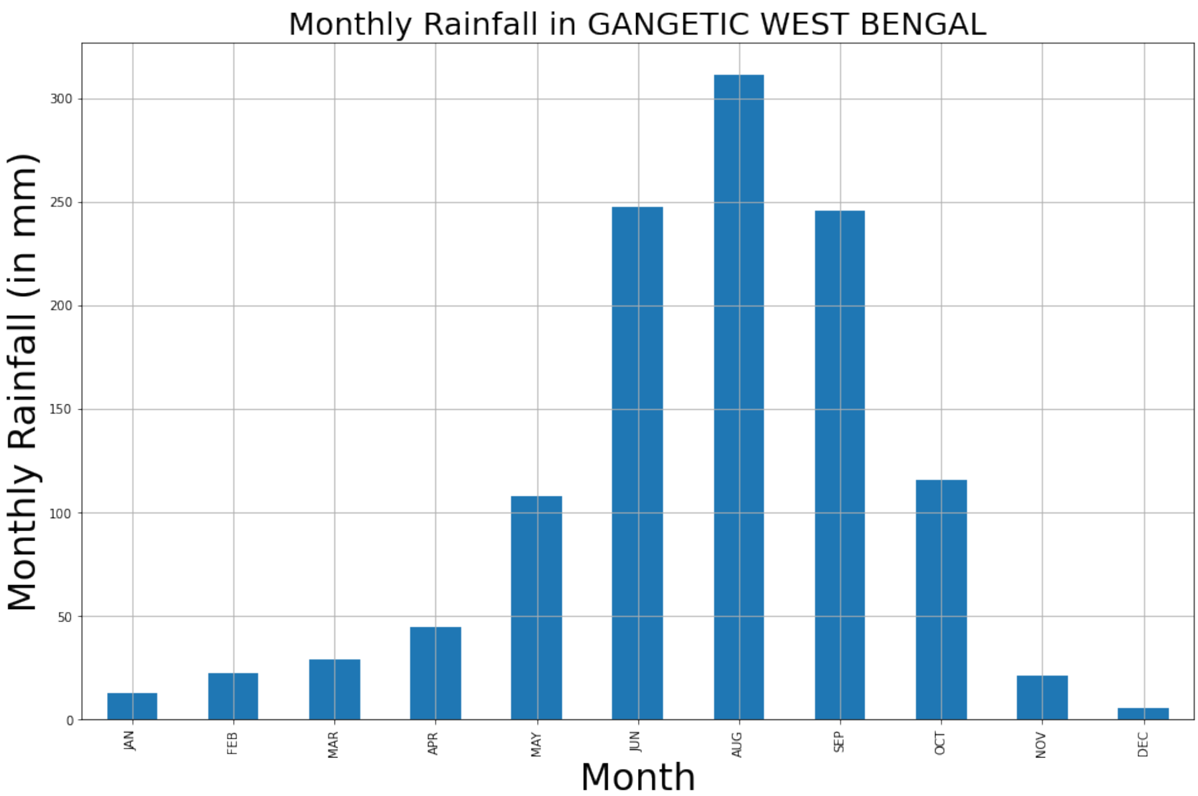


Figure 3-3: Average Monthly rainfall in the Gangetic West Bengal region

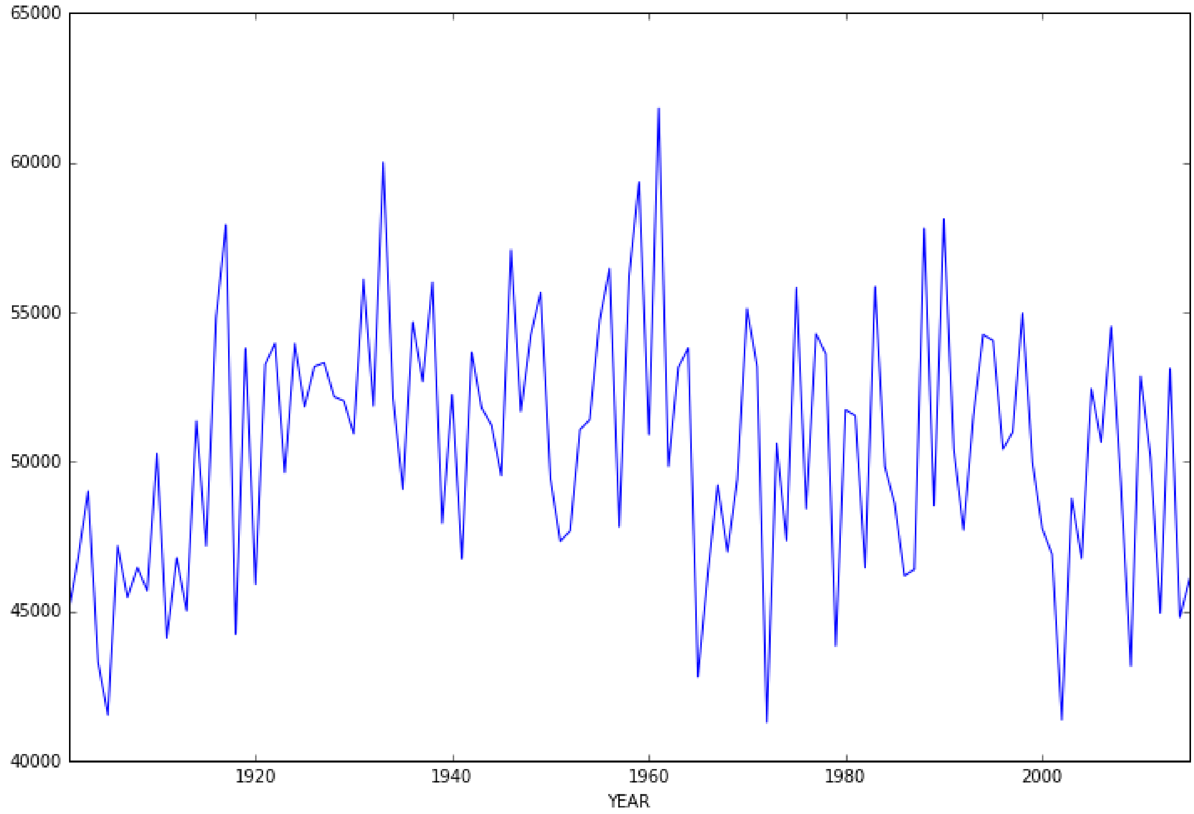


Figure 3-4: Distribution of rainfall over years (Highest in 1950)

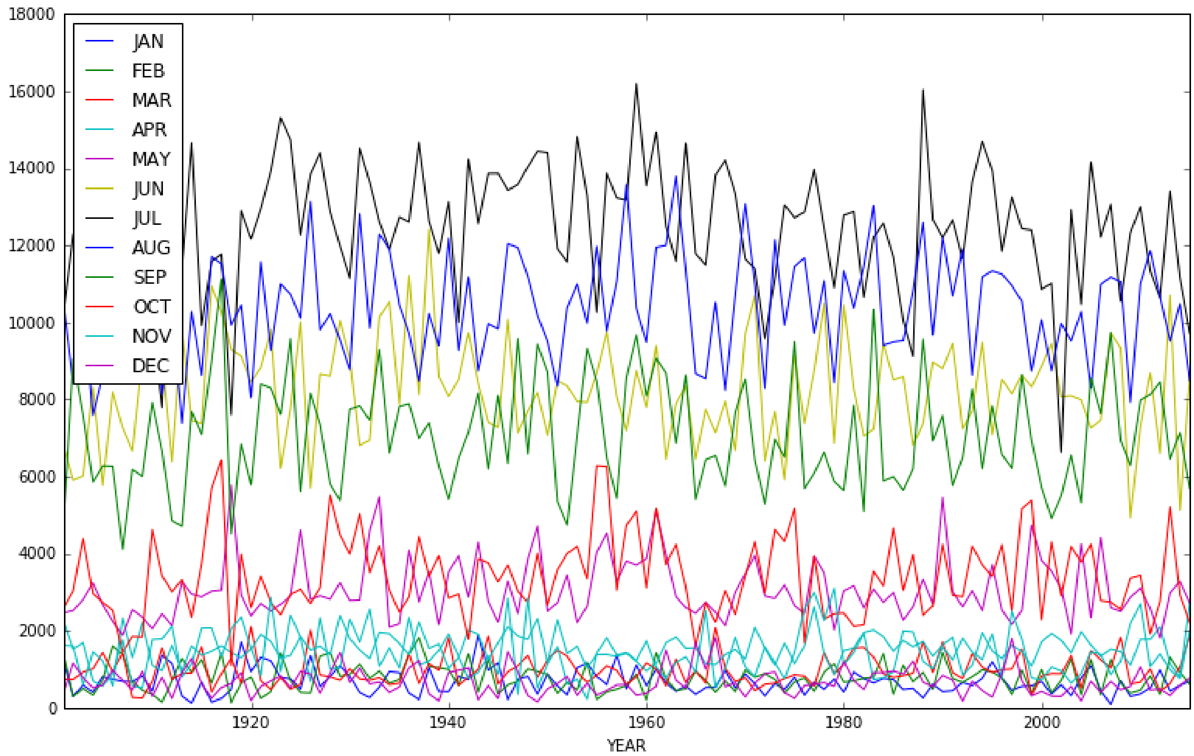


Figure 3-5: The distribution of rainfall over months

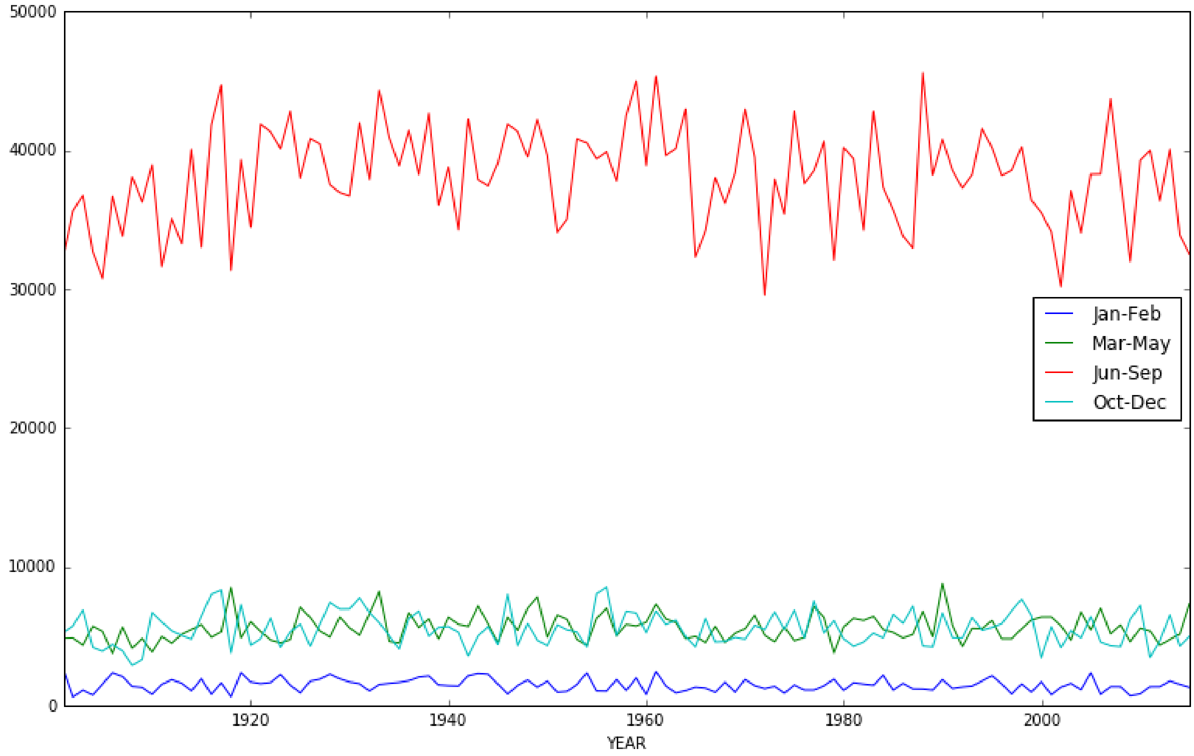
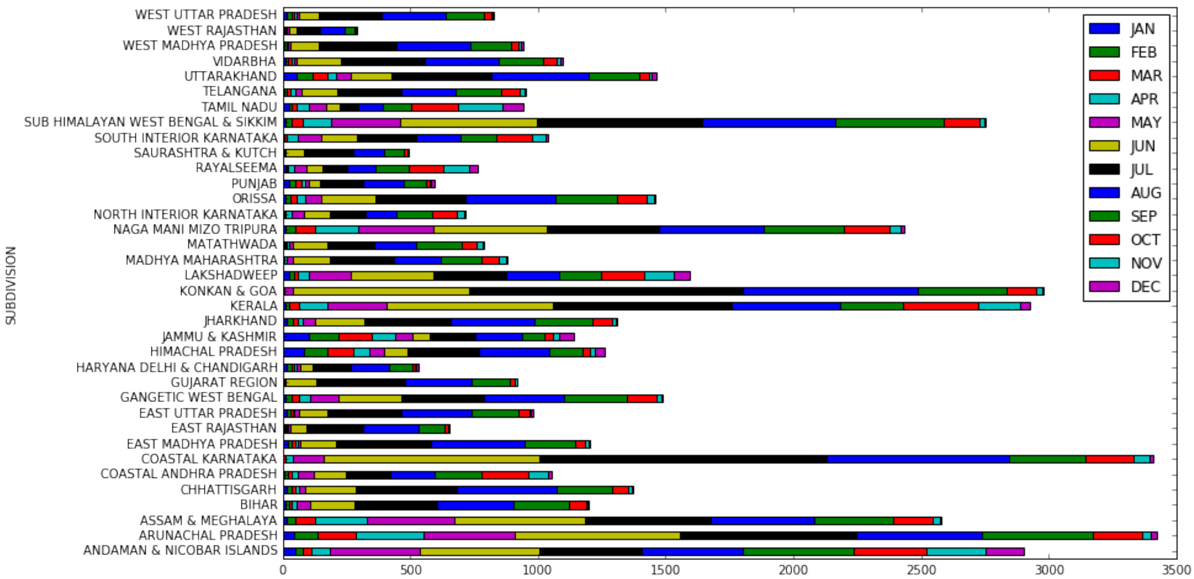


Figure 3-6: The distribution of rainfall for the months grouped together (quarterly)



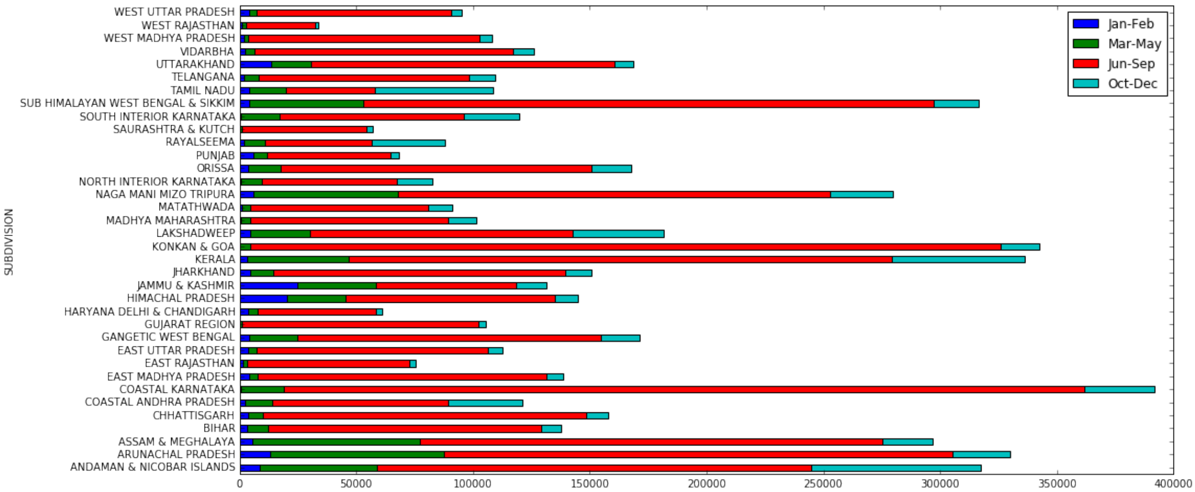
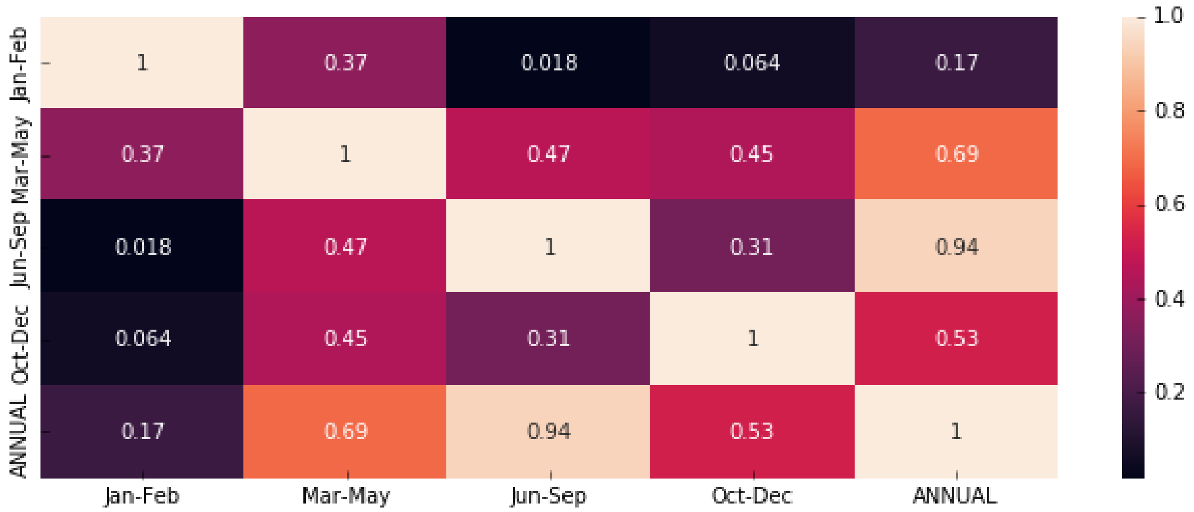


Figure 3-7: The two graphs show that the amount of rainfall is reasonably good in the months of march, april, may in eastern India



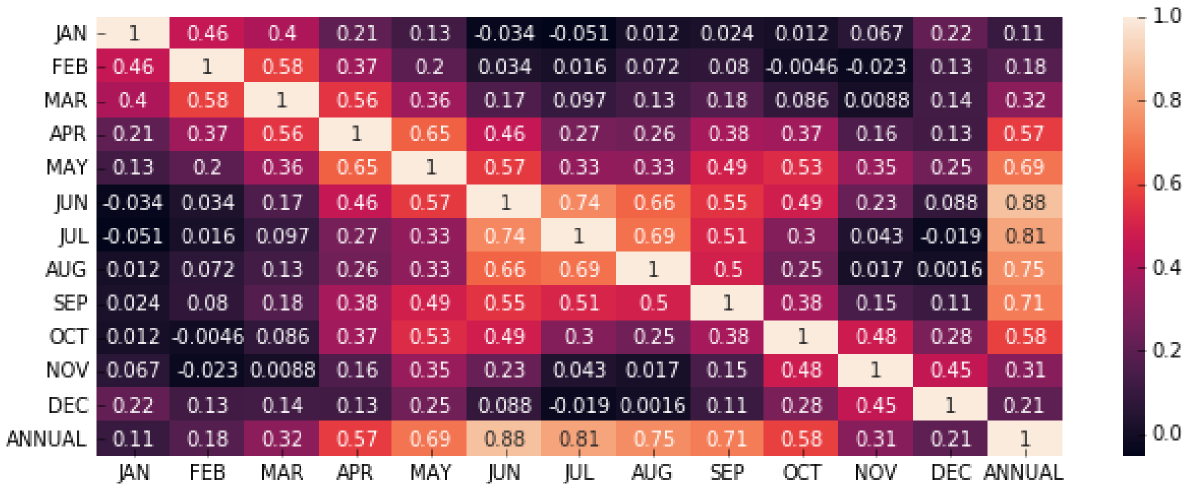


Figure 3-8: Heat Map showing the co-relation (dependency) between the amounts of rainfall over months.

Methods & Results

With this data with more variations of average rainfall, it is very difficult for a statistical model to predict the required data point. Here we implement neural networks to predict the avg rainfall, the neural net is used to create multiple features that helps in predicting the data points with more seasonal variations.

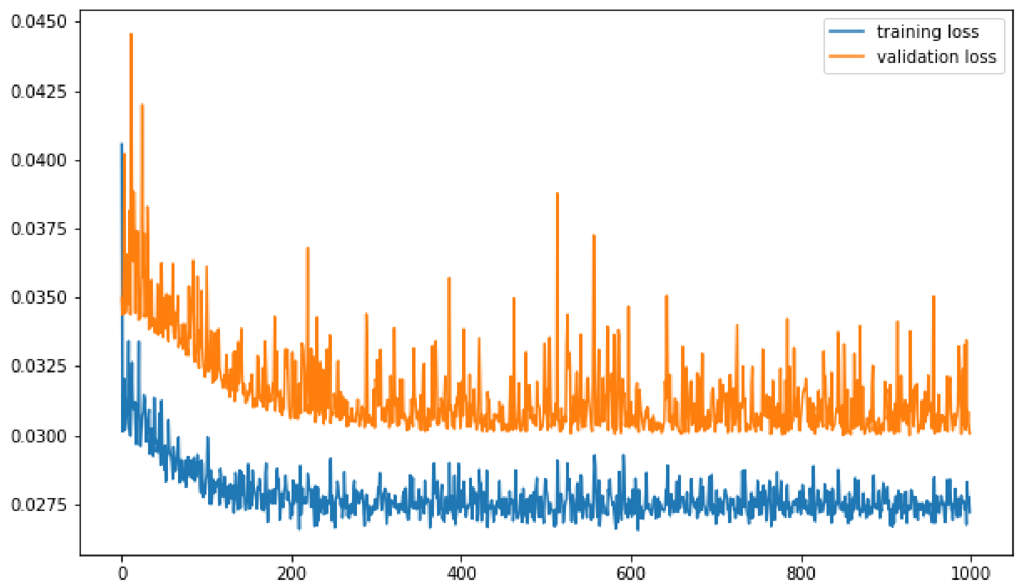


Figure 4-1: Plot of Loss vs Number of epochs

From the plot, we can see that the model has comparable performance on both train and validation data sets as the difference on train set and validation set is less and becomes constant after a certain number of epochs. We can also use methods like Early Stopping call back function if the performance becomes stagnant on increasing the number of epochs.

Long Short-Term Memory (LSTM) is a type of recurrent neural network that can learn the order dependence between items in a sequence. LSTMs have the promise of being able to learn the context required to make predictions in time series forecasting problems, rather than having this context pre-specified and fixed.

The intuition behind using LSTM is that these models are pretty good at extracting patterns in input feature space, where the input data spans over long sequences. Also, unlike regression predictive modeling, time series also adds the complexity of a sequence dependence among the input variables. Given the gated architecture of LSTM's that has this ability to manipulate its memory state, they are ideal for such problems. This adds a great benefit in the problems involving time series forecasting, where the classical linear methods can be difficult to adapt to multivariate or multiple input forecasting problems.

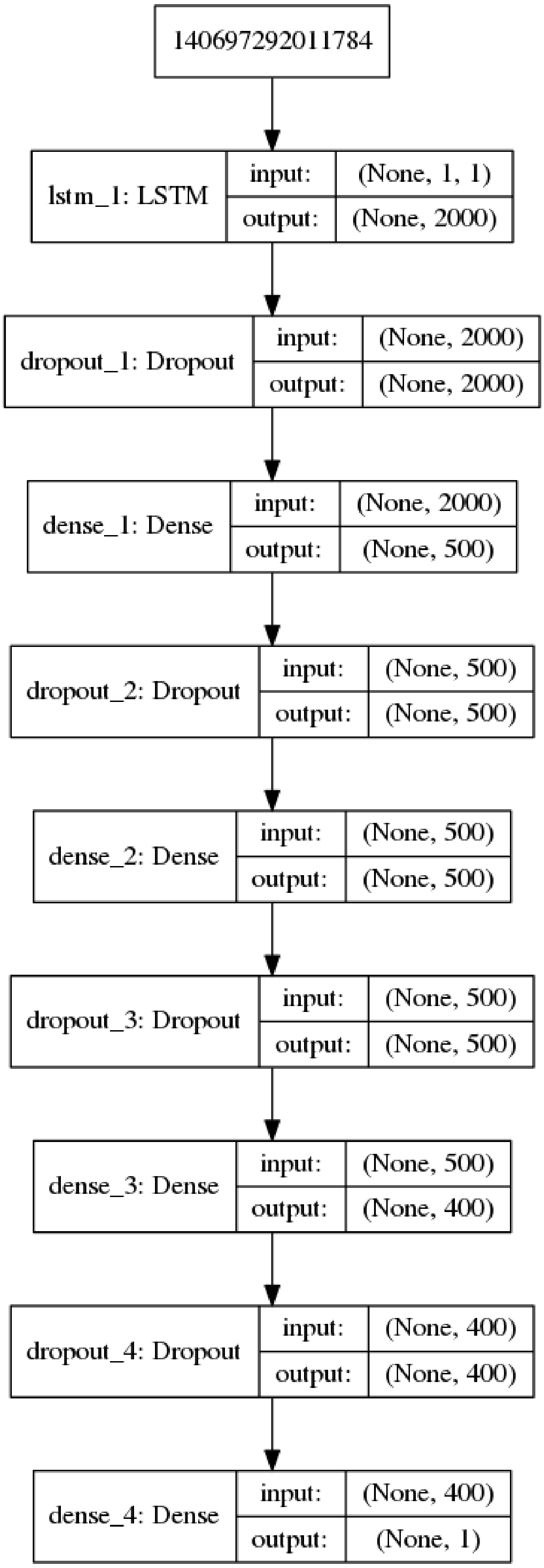


Figure 4-2: Model architecture

4.1 Extreme Value Analysis

We will fit the predicted data to GEV distribution for extreme analysis. In probability theory and statistics, the generalized extreme value (GEV) distribution is a family of continuous probability distributions developed within extreme value theory to combine the Gumbel, Fréchet and Weibull families also known as type I, II and III extreme value distributions.

The generalized extreme value distribution is defined by the following distribution function

$$F(x) = e^{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}},$$

for $1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0$, ξ the shape parameter, μ the location parameter and $\sigma > 0$ the scale parameter. We can derive a density function

$$f(x) = \frac{1}{\sigma}\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}-1} e^{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}}.$$

The parameters(scale, location and shape are plotted in the histogram in Figure 4-3

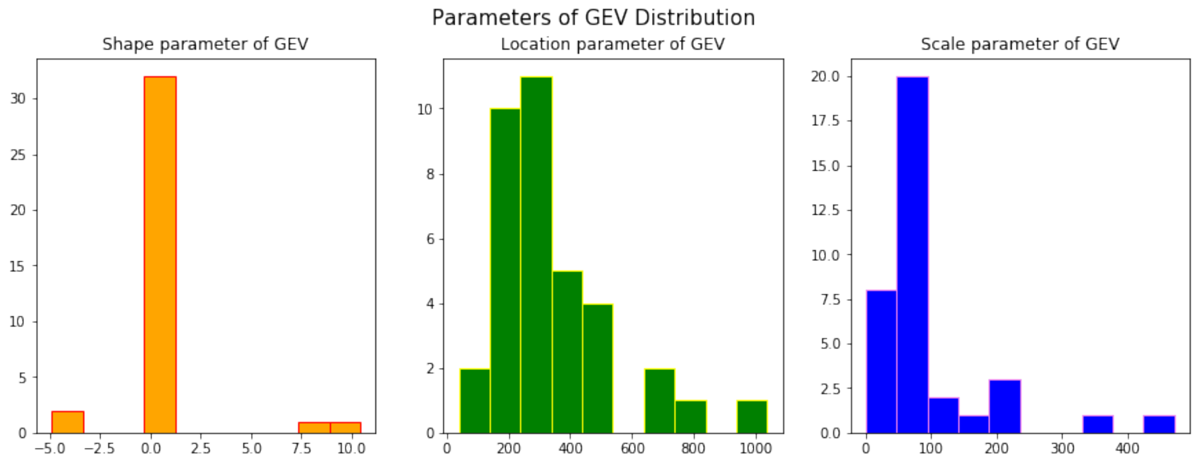


Figure 4-3: The histogram of GEV parameters

From the fitted distribution, we can estimate how often the extreme quantiles occur with a certain return level. The return value is defined as a value that is expected to be equaled or exceeded on average once every interval of time (T) (with a probability of 1/T).

A general trend of return levels from the distribution is plotted against return period in Figure 4-4 for all the subdivision (location) in India.

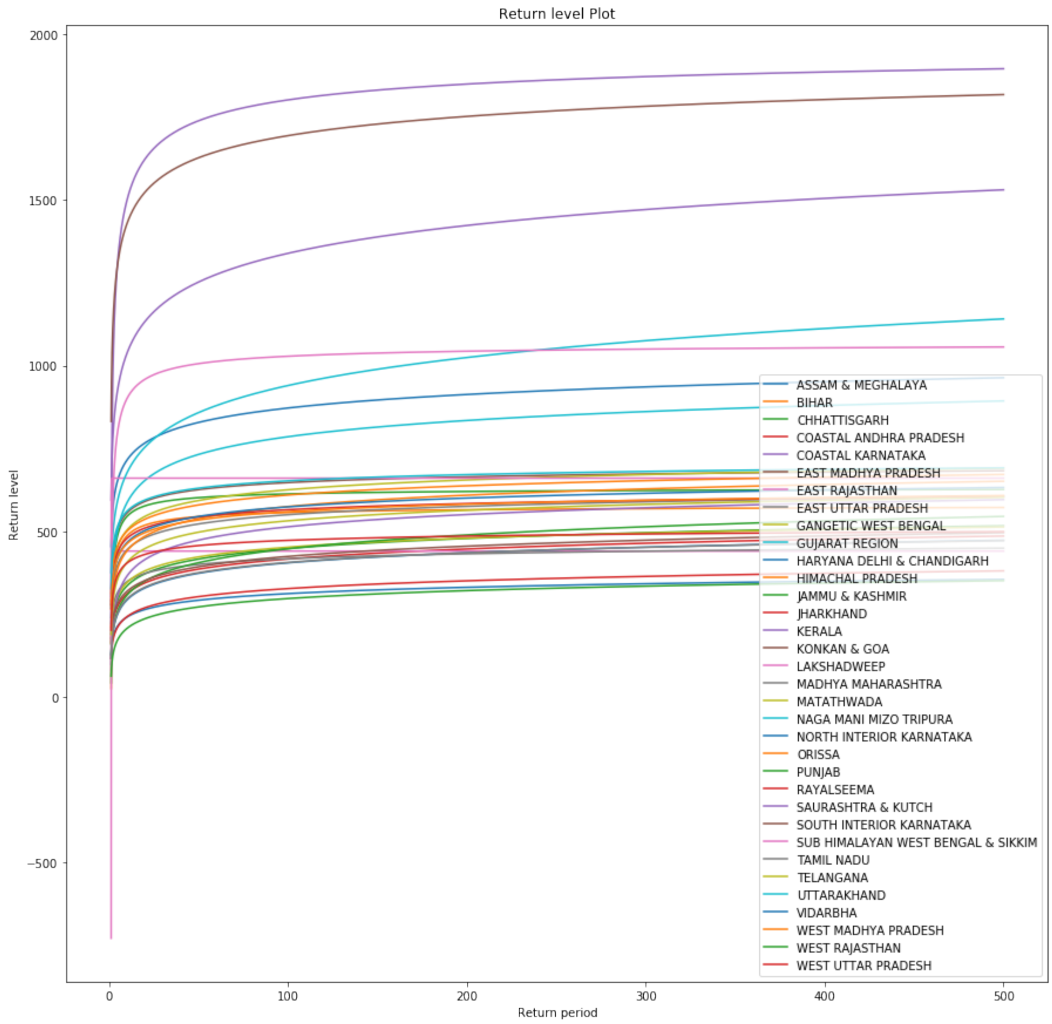


Figure 4-4: Return Level plot for various sub divisions in India

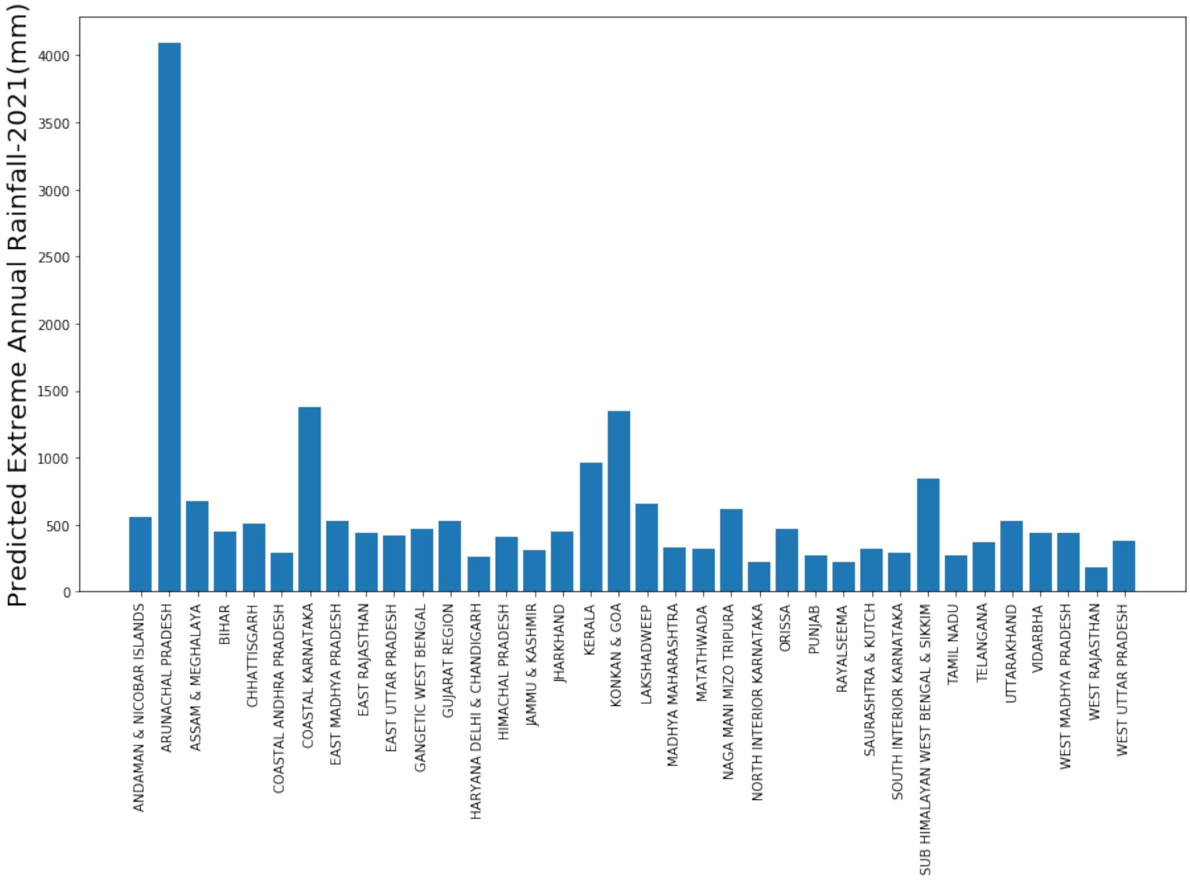


Figure 4-5: Predicted extreme Annual Rainfall for various subdivision for 2021

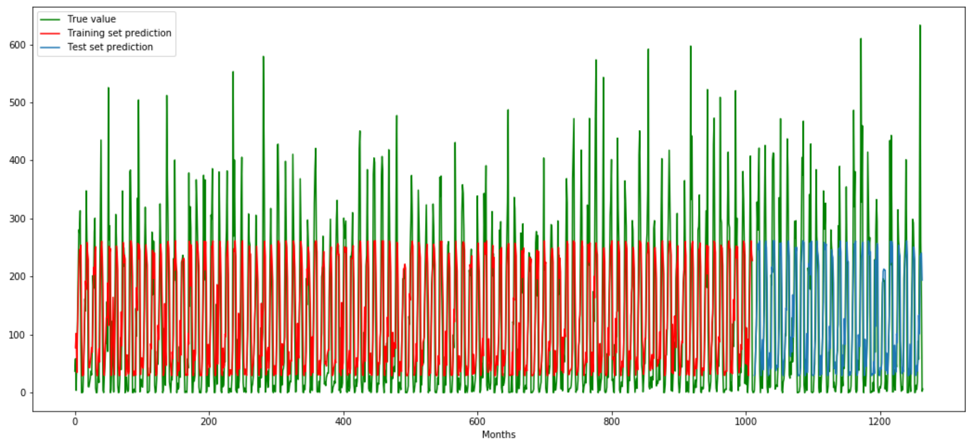


Figure 4-6: Training set & Test set prediction with True value

References

Bhowmik RS, D. V. (2010). Application of multimodel ensemble techniques for real time district level rainfall forecasts in short range time scale over indian region. *Meteorol atmos phys 106:19–35*.

G. Xu, V. C. (2005). Operational feasibility of neural-network-based radar rainfall estimation. *Ieee geoscience and remote sensing letters, 2 (1) (2005), pp. 13-17*.

H. Liu, G. X., V. Chandrasekar. (2001). An adaptive neural network scheme for rainfall estimation from WSR-88d observations. *Journal of applied meteorology, 40 (11) (2001), pp. 2038-2050*.

Hart RE, G. R. (2001). Using normalized climatological anomalies to rank synoptic-scale events objectively. *Mon weather rev 129:2426–2442*.

JC, S. (2004). Finite difference schemes and partial differential equations. *Siam: Society for industrial and applied mathematics, philadelphia 435 pp*.

Khaladkar RM, M. P., Narkhedkar SG. (2007). Performance of NCMRWF models in predicting high rainfall spells during SW monsoon season—a study for some cases in july 2004. *Report no. rr-116*.

Nayak, P., Sudheer, K., Rangan, D., & Ramasastri, K. (2004). A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, *291*(1), 52–66. https://doi.org/https://doi.org/10.1016/j.jhydrol.2003.12.010

R. Aliev, F. A., K.W. Bonfig. (2000). Soft computing: Eine grundlegende einführung verlag technik, berlin (2000).

Sahai AK, S. V., Soman MK. (2000). All india summer monsoon rainfall prediction using an artificial neural network. *Clim dyn 16:291–302*.

Teschl, R., Randeu, W. L., & Teschl, F. (2007). Improving weather radar estimates of rainfall using feed-forward neural networks [Computational Intelligence in Earth and Environmental Sciences]. *Neural Networks*, *20*(4), 519–527. https://doi.org/https://doi.org/10.1016/j.neunet.2007.04.005