**GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY**

**PROJECT TITLE: Rainfall Prediction**

**MINOR PROJECT**

**ICT-497**

**Submitted in partial fulfilment of the requirements**

**for the award of the degree of**

**Bachelor of Technology**

**in**

**Computer Science Engineering**

**at**

**UNIVERSITY SCHOOL OF INFORMATION, COMMUNICATION AND TECHNOLOGY**

**Supervisor: Dr. Jyotsana Yadav Submitted By: Bharat Sahay**

**Enrollment No: 01816403221**

**Programme: B.Tech. CSE 7th Sem**

**Table of Contents**

1. Declaration
2. Certificate of Originality
3. Introduction
4. Objectives
5. Technology Used
6. Work Done
7. Planned Features
8. References

**DECLARATION**

I, **Karan Chhillar**, a student of Bachelor of Technology in Computer Science & Engineering at the **University School of Information, Communication & Technology (USICT), Guru Gobind Singh Indraprastha University, New Delhi**, hereby declare that the work presented in this minor project report titled **"Rainfall Prediction using Machine Learning"** is an original and authentic effort undertaken by me under the technical guidance of **Prof. Amit Prakash, USICT**.

I affirm that due acknowledgment has been made wherever necessary, and the information used in this project has been properly cited and referenced. The progress report has been checked by Turnitin Plagiarism Detection software at GGSIPU. The work is free from any sort of plagiarism, fabrication, falsification, or copyright issues, and the similarity of the contents identified is as per UGC norms.

I declare that the work in this project has not been submitted in full or in part for any diploma or degree course of this University or any other University/Institute to the best of my knowledge and belief. I shall be solely responsible for any copyright infringement or plagiarism, if any, in the said project or any dispute arising out of my work. My supervisor or anyone else shall not be held responsible for any full or partial violation of copyright/intellectual property rights/plagiarism issues. All the inputs and suggestions by the supervisor and SRC have been incorporated into the project report.

Karan Chhillar  
B.Tech CSE Student

(Enrollment No: 05716403221)  
University School of Information, Communication & Technology,  
Guru Gobind Singh Indraprastha University,  
Sector-16C, Dwarka, New Delhi, India-110078

Date-25/10/2024

**UNIVERSITY SCHOOL OF INFORMATION, COMMUNICATION AND TECHNOLOGY**

**GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY**

****

**Certificate of Originality**

This is to certify that the work embodied in this project report titled **"Rainfall Prediction using Machine Learning"** is an original piece of work carried out by **Karan Chhillar**, a student of Bachelor of Technology in Computer Science & Engineering at the **University School of Information, Communication & Technology (USICT), Guru Gobind Singh Indraprastha University**. This project has been completed under the supervision of **Prof. Amit Prakash, USICT**.

The work has not been submitted, in part or in full, for any other degree or diploma at this or any other university, based on the declaration of the candidate.

It is certified that the work being submitted for this project report is free from any form of falsification, fabrication of results, data, claims, copyrights, and plagiarism, to the best of my knowledge and belief.

**Prof. Amit Prakash**  **Karan Chhillar**Supervisor B.Tech CSE Student  
University School of Information, Communication & Technology 05716403221  
Guru Gobind Singh Indraprastha University,  
New Delhi

Date-

**Abstract**

This project focuses on developing a machine learning model for **rainfall prediction** using the **Indian Meteorological Department's (IMD) Yearly Gridded Rainfall dataset**. Accurate rainfall prediction is crucial for weather forecasting, agricultural planning, and disaster management, as it helps mitigate risks associated with natural disasters such as floods and droughts. The project aims to enhance the accuracy of weather predictions by leveraging machine learning techniques to analyze historical rainfall data and forecast future patterns.

The project is implemented using a variety of tools and technologies, including **Python**, **Pandas**, **NumPy**, **Scikit-learn**, and **Matplotlib**, to ensure efficient data preprocessing, modeling, and visualization. Key steps include data preprocessing, exploratory data analysis (EDA), model selection, and performance evaluation. Several machine learning algorithms, such as **Linear Regression**, **Random Forest**, and **Neural Networks**, are applied and evaluated based on performance metrics like **Mean Squared Error (MSE)** and **R-squared**.

This report outlines the entire development process, detailing the machine learning models used, the challenges encountered, and the effectiveness of the predictions made. The project aims to provide actionable insights for weather forecasting and climate research, contributing to improved decision-making in sectors like agriculture and disaster management.

**Introduction**

Accurate rainfall prediction plays a crucial role in various sectors such as weather forecasting, agriculture, and disaster management. It helps mitigate the adverse effects of natural disasters like floods and droughts, while also aiding in efficient agricultural planning. This project, titled **"Rainfall Prediction using Machine Learning,"** is designed to address the need for precise rainfall forecasts by developing a machine learning model that can predict future rainfall patterns based on historical data from the **Indian Meteorological Department's (IMD) Yearly Gridded Rainfall dataset**.

The core objective of this project is to leverage machine learning techniques to analyze historical rainfall data, identify patterns, and produce accurate predictions. The dataset consists of yearly rainfall records across different regions of India, presented in a grid format. This allows the model to capture geographical variations and trends over time, providing valuable insights into future rainfall distribution.

The project involves several critical steps, including data preprocessing, exploratory data analysis (EDA), model selection, and performance evaluation. Various machine learning algorithms such as Linear Regression, and various ML algorithms are employed to build the predictive model. The project also emphasizes the importance of data visualization, with graphs and heatmaps used to compare predicted rainfall values with actual outcomes.

This report details the methodologies used in the project, the technologies applied, and the results achieved, while also outlining future enhancements to improve the model's accuracy and reliability. By combining simplicity in implementation with powerful predictive capabilities, this project aims to contribute to more effective weather forecasting and climate research, benefiting sectors that rely on accurate rainfall predictions.

**Technology Used**

The **Rainfall Prediction using Machine Learning** project employs a range of modern tools and technologies to build an efficient and accurate machine learning model for rainfall forecasting. The project is developed using **Python** due to its extensive library support and ease of use in data processing and model development. The following technologies and libraries were utilized:

1. **Pandas**: This powerful data manipulation library is used to handle and preprocess the dataset. Pandas enables efficient cleaning, normalization, and transformation of the rainfall data, making it ready for modeling.
2. **NumPy**: NumPy provides support for high-performance numerical computations. It is used in this project for efficient handling of large datasets and performing mathematical operations, which are crucial in the model training process.
3. **Scikit-learn**: Scikit-learn is a machine learning library that simplifies the implementation of various algorithms. In this project, models such as **Linear Regression**, **Random Forest**, and **Neural Networks** are implemented using Scikit-learn. It also offers tools for model evaluation through metrics like **Mean Squared Error (MSE)** and **R-squared**.
4. **Matplotlib**: For data visualization, Matplotlib is used to generate insightful graphs and heatmaps. These visualizations help compare predicted and actual rainfall values, providing clear insights into model performance and future trends.
5. **Seaborn**: In addition to Matplotlib, Seaborn is employed for advanced visualization tasks, such as generating correlation matrices and detailed heatmaps, making exploratory data analysis (EDA) more intuitive.
6. **Jupyter Notebook**: This interactive environment is used for the development and execution of the project code. It allows for easy iteration during the model development process, enabling real-time visualization of results and facilitating rapid prototyping.

By integrating these technologies, the project ensures a robust and scalable solution for rainfall prediction, making the most of advanced machine learning techniques and effective data handling practices. The combination of these tools enables the creation of an accurate model that can assist in weather forecasting and agricultural planning.

**Work Done**

The development of the **Rainfall Prediction using Machine Learning** project has primarily involved data preprocessing, exploratory data analysis, and model selection. Below is a summary of the work completed thus far:

**1. Data Preprocessing:**

* Cleaned the dataset by handling missing values to ensure the data was complete and reliable for model training.
* Normalized the dataset to standardize the range of rainfall values, making it suitable for machine learning algorithms.
* Split the data into training and testing sets for accurate model evaluation.

**2. Exploratory Data Analysis (EDA):**

* Conducted an in-depth analysis of the historical rainfall data to identify patterns, trends, and seasonality.
* Visualized geographical distribution of rainfall using heatmaps and identified correlations between rainfall and different regions of India.
* Summarized rainfall trends over the years to understand the data better before applying machine learning models.

**3. Model Development:**

* Implemented several machine learning algorithms, including Linear Regression, Random Forest, and Neural Networks, to predict future rainfall based on the historical dataset.
* Trained each model on the processed data, comparing performance across various algorithms.
* Focused on tuning hyperparameters to enhance model accuracy.

**4. Model Evaluation:**

* Evaluated the performance of the models using metrics such as Mean Squared Error (MSE) and R-squared (R²) to determine their predictive power.
* Visualized the predicted versus actual rainfall to gauge the effectiveness of the models and to pinpoint areas for improvement.

This work has established a strong foundation for further model refinement and real-time rainfall prediction, ensuring the project is on track for successful completion.

**Dataset Description:**

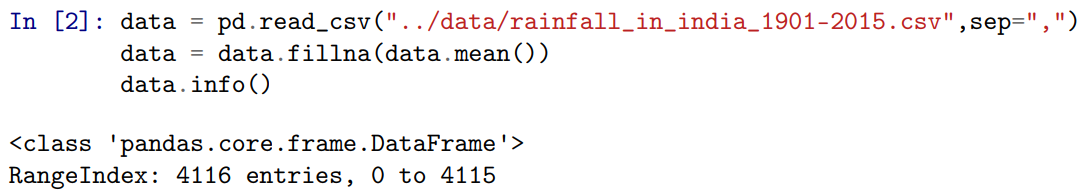
**Dataset link =** https://www.kaggle.com/datasets/rajanand/rainfall-in-india

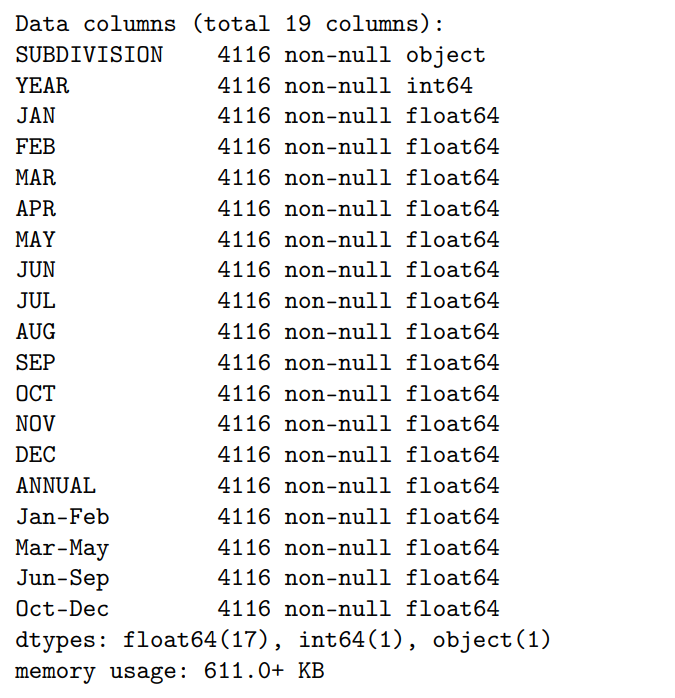
• Data has 36 sub divisions and 19 attributes (individual months, annual, combinations of 3 consecutive

months).

• For some of the subdivisions data is from 1950 to 2015.

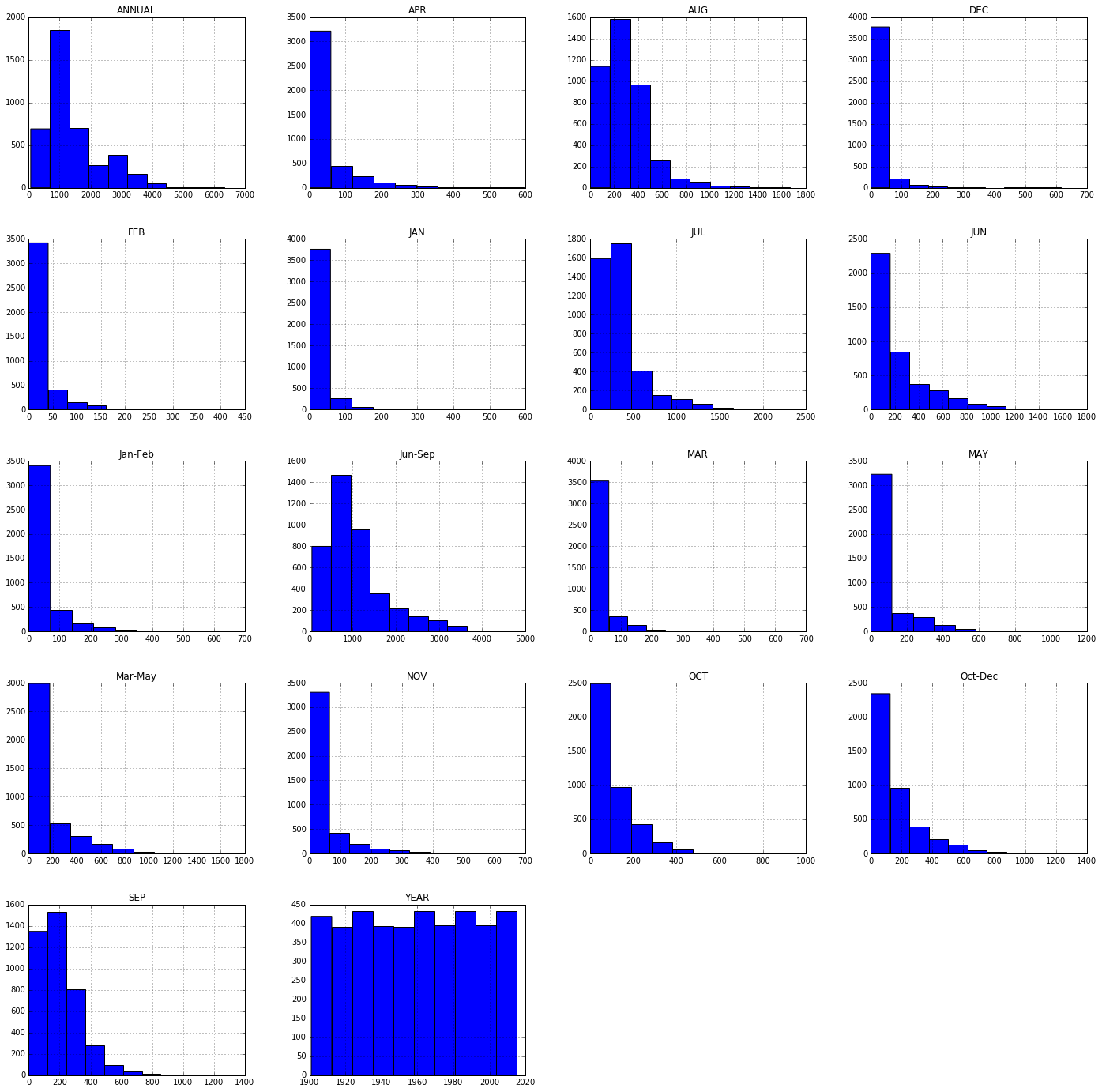
• All the attributes has the sum of amount of rainfall in mm.

****

****

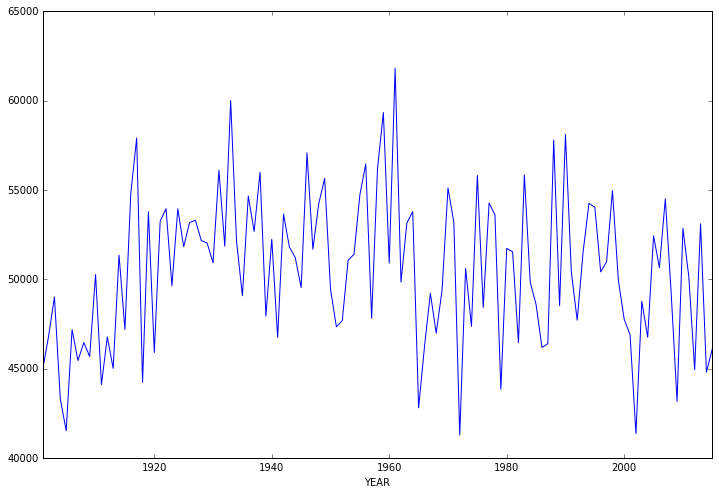
**Observations:**

****

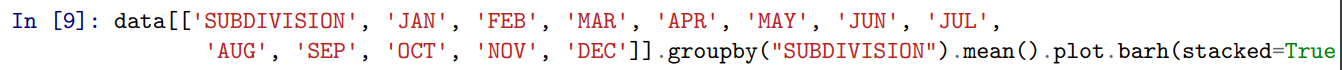


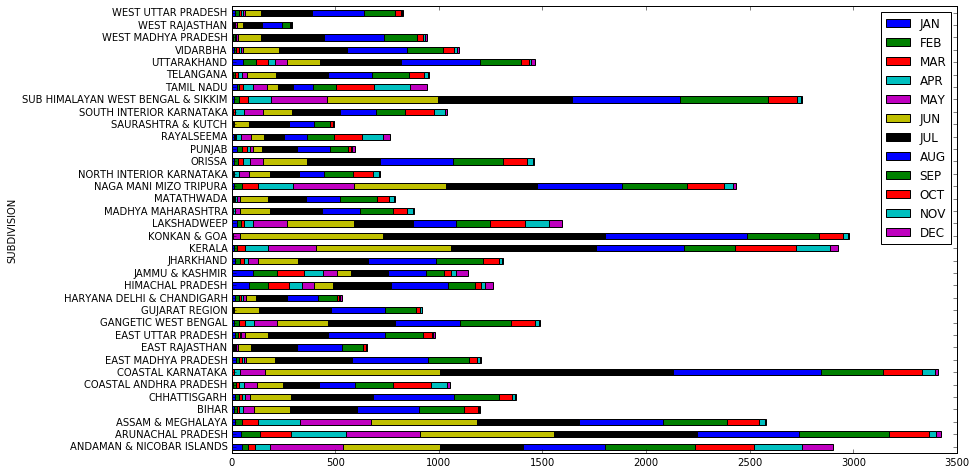
• Above histograms show the distribution of rainfall over months.

• Observed increase in amount of rainfall over months July, August, September.

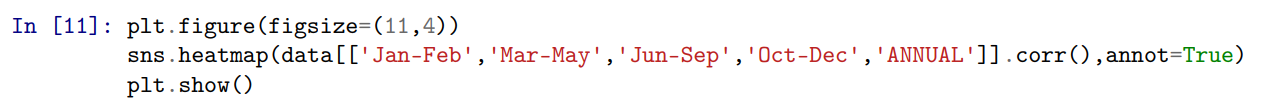


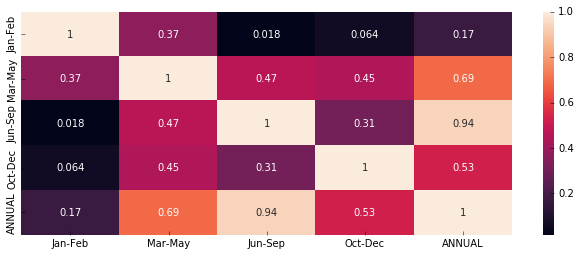
• The above graph shows the distribution of rainfall over months.



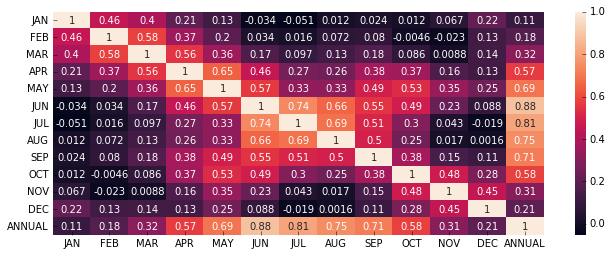


• Above graph shows that the amount of rainfall is reasonably good in the eastern India.



****

****

****

• Heat Map shows the co-relation(dependency) between the amounts of rainfall over months.

• From above it is clear that if 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 b good in the overall year.

**Predictions:**

• For prediction we formatted data in the way, given the rainfall in the last three months we try to

predict the rainfall in the next consecutive month.

• For all the experiments we used 80:20 training and test ratio.

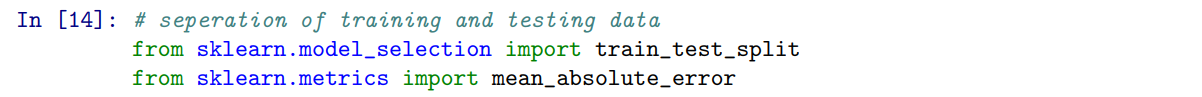
1. Linear regression
2. SVR
3. Artificial neural nets

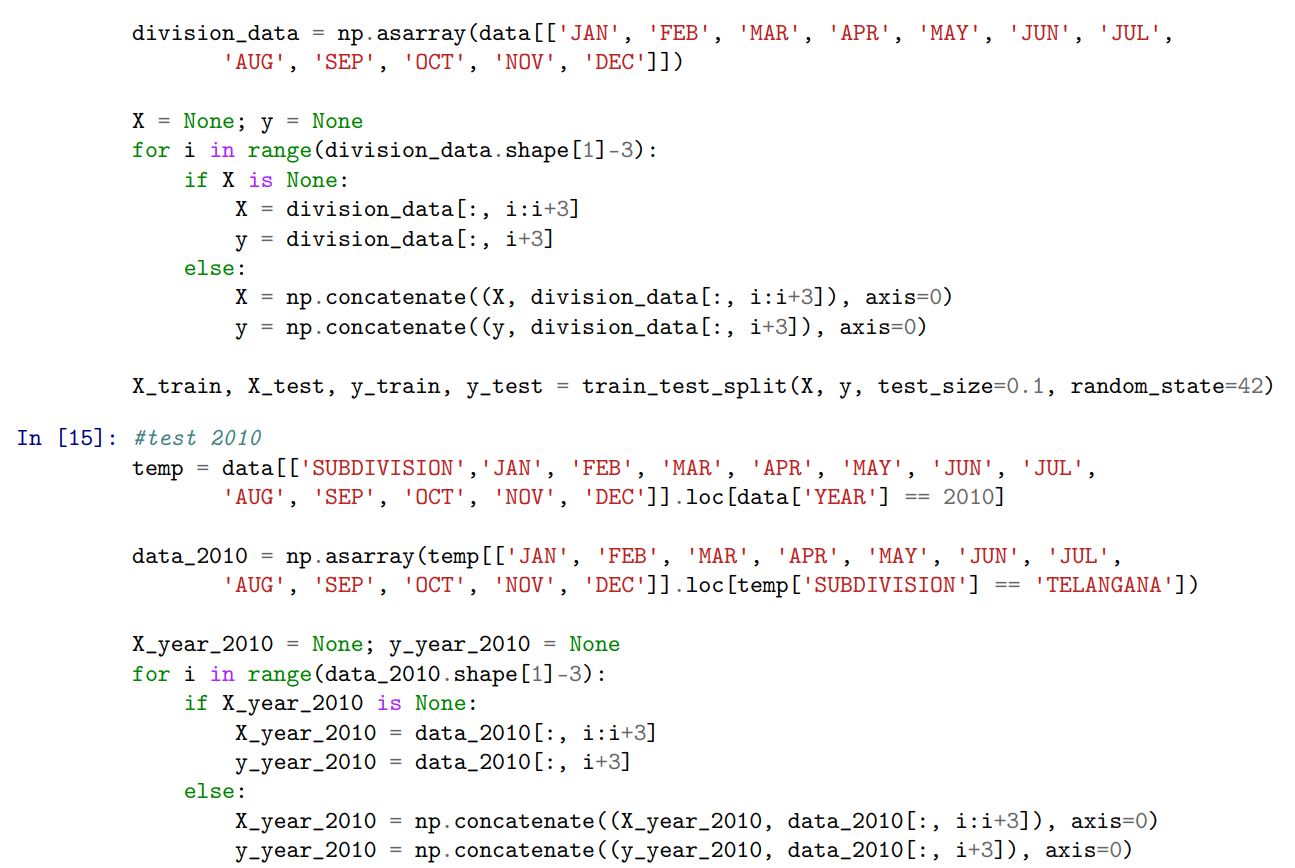
• Testing metrics: We used Mean absolute error to train the models.

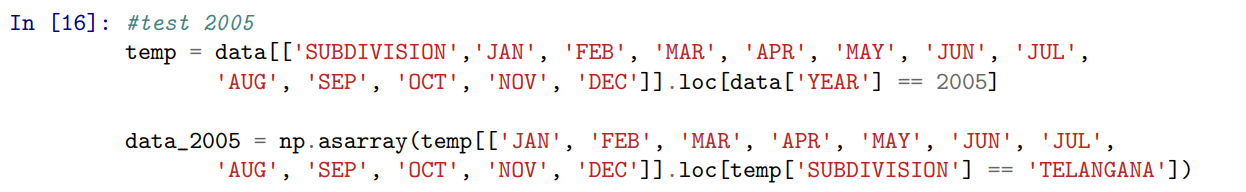
• We also shown the amount of rainfall actually and predicted with the histogram plots.

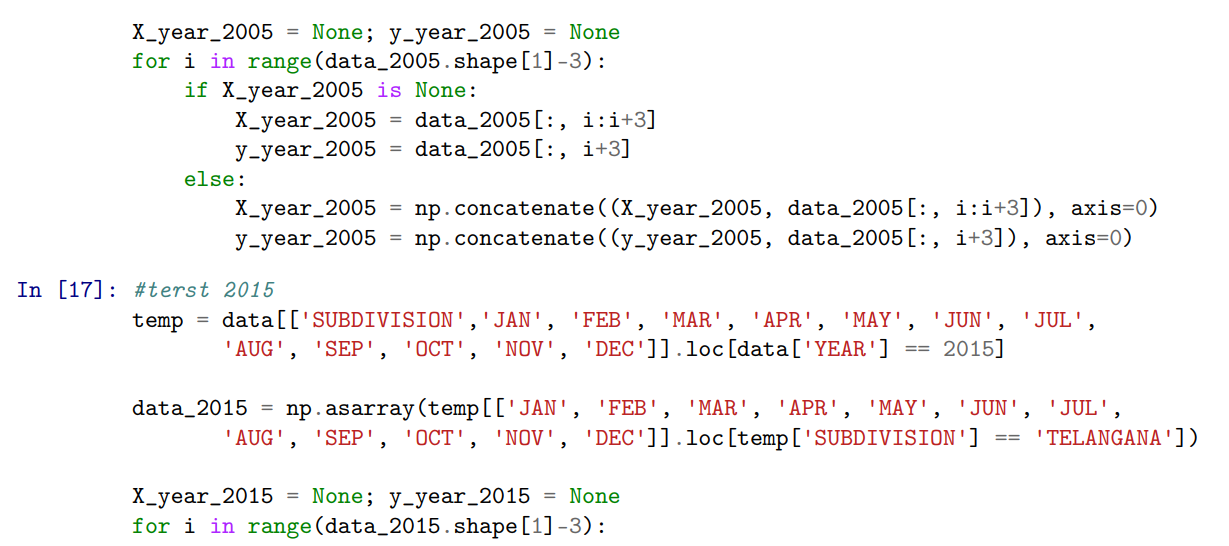
• We did two types of trainings once training on complete dataset and other with training with only telangana data

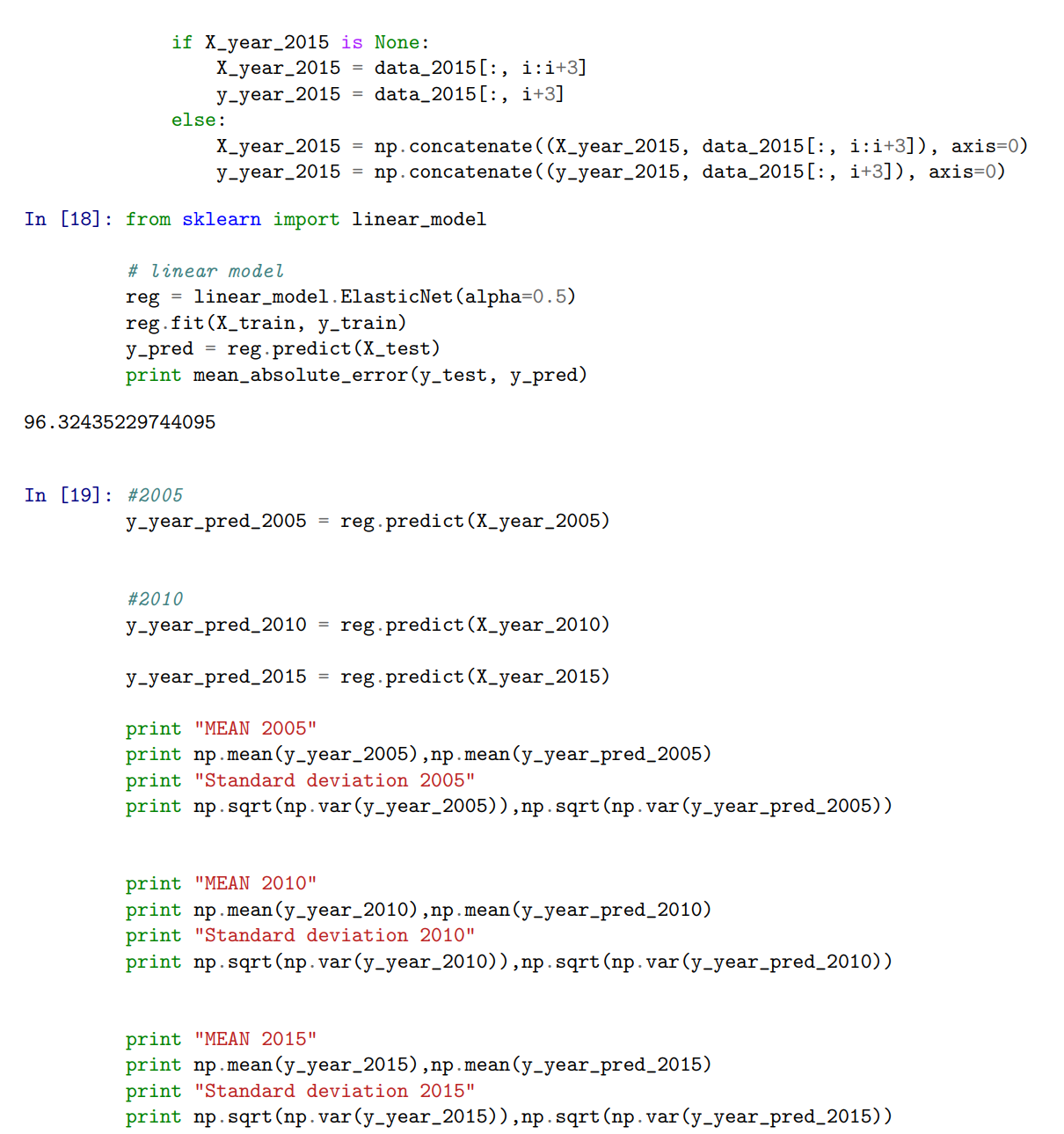
• All means are standard deviation observations are written, first one represents ground truth, second one represents predictions.

****

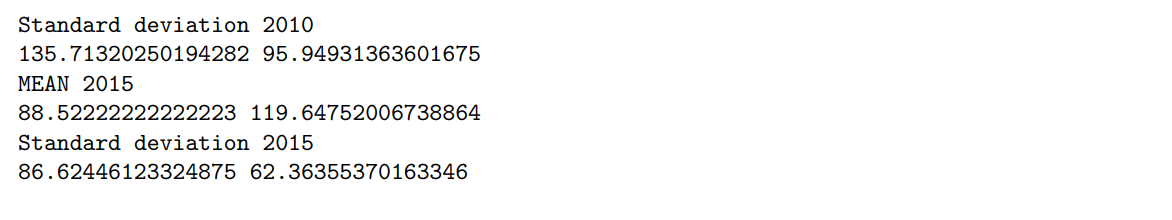
****

****

****

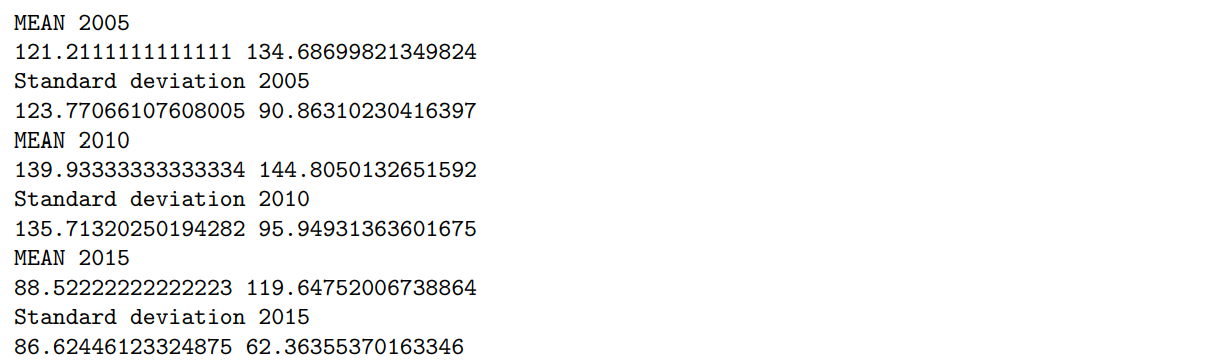
****

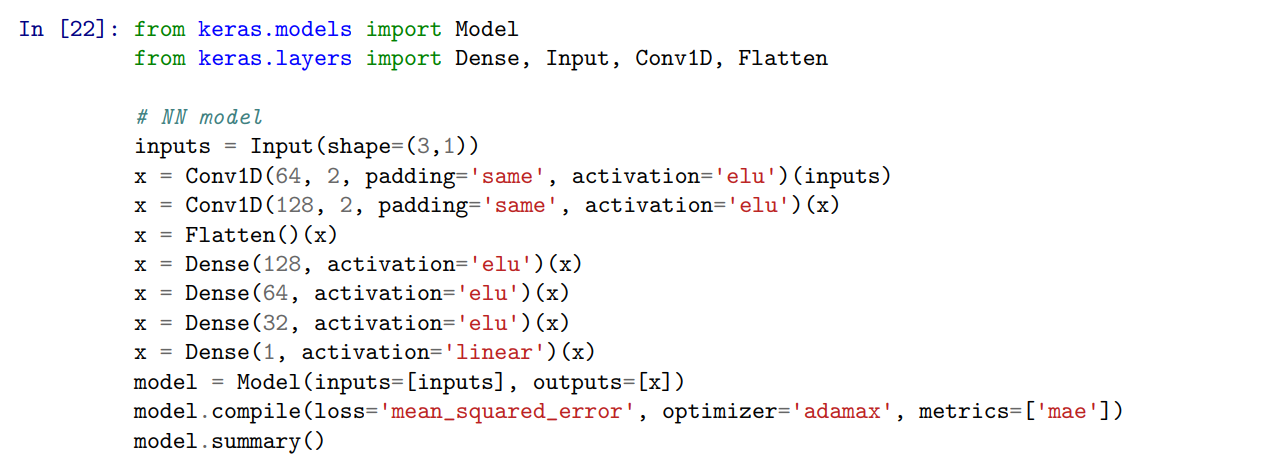
****

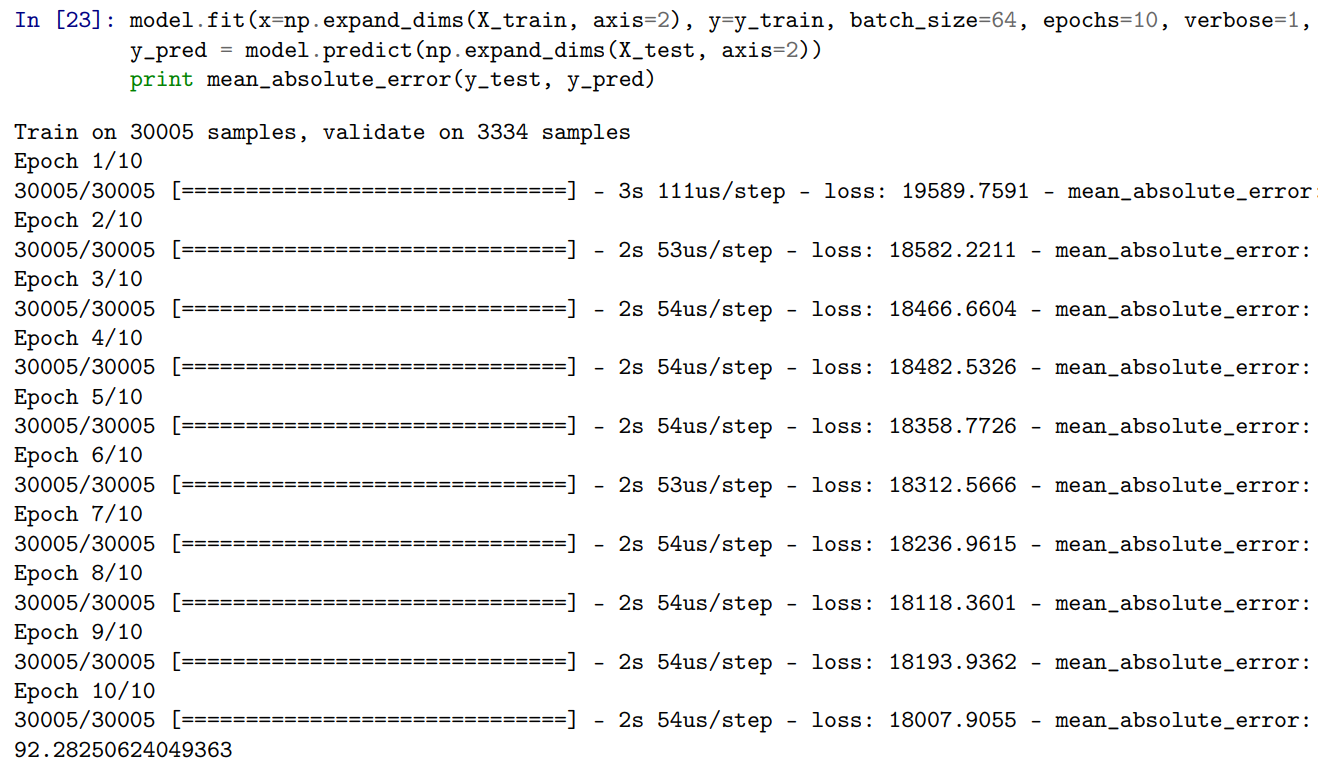
****

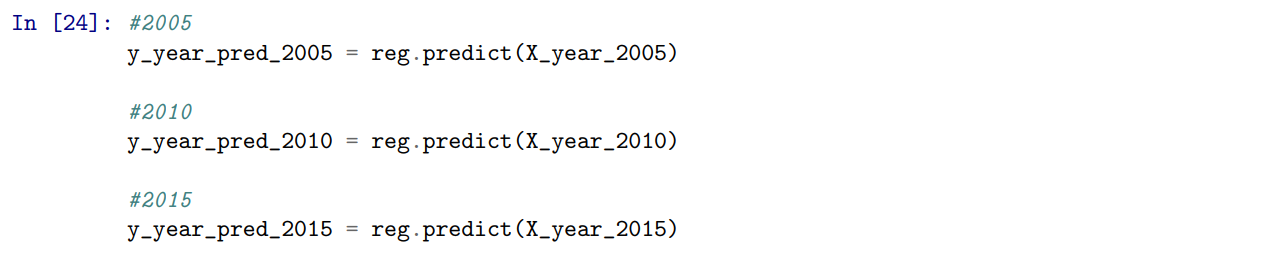
****

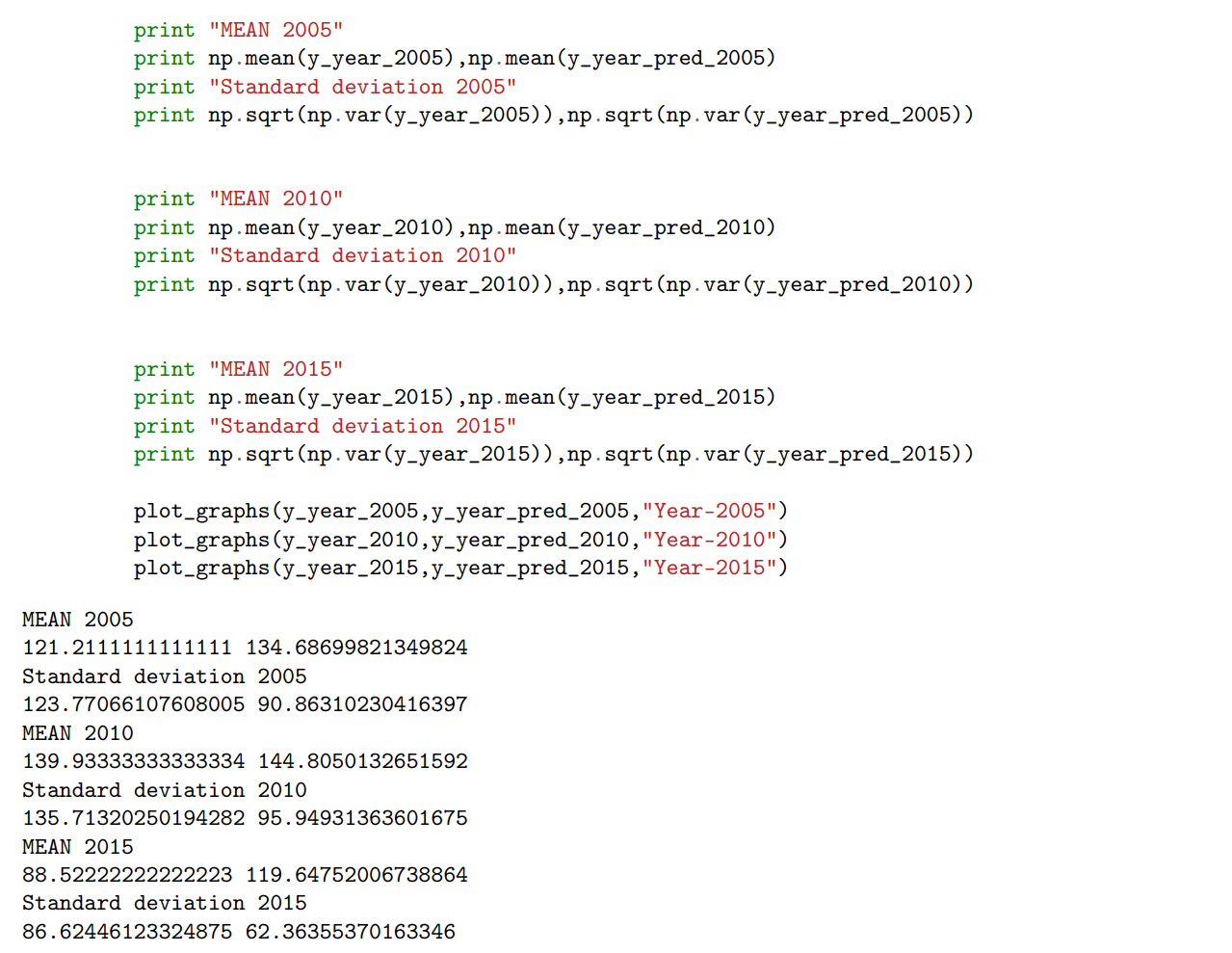
****

****

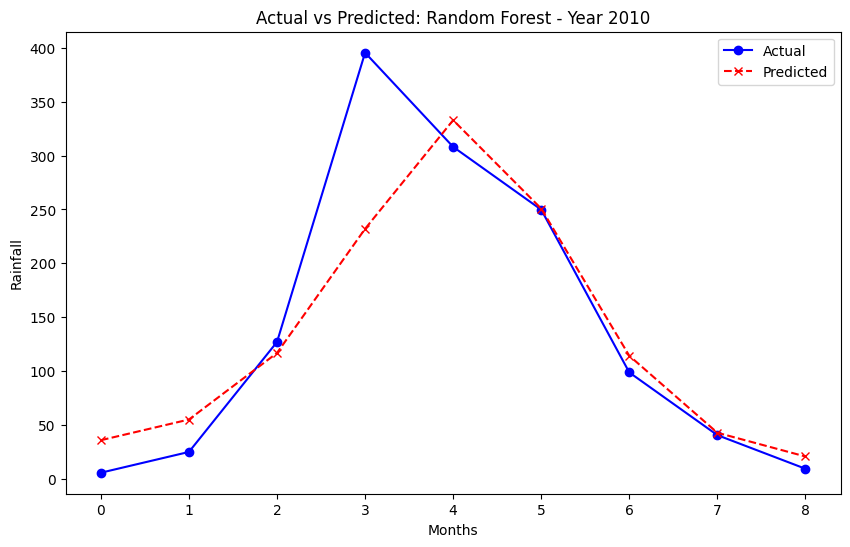
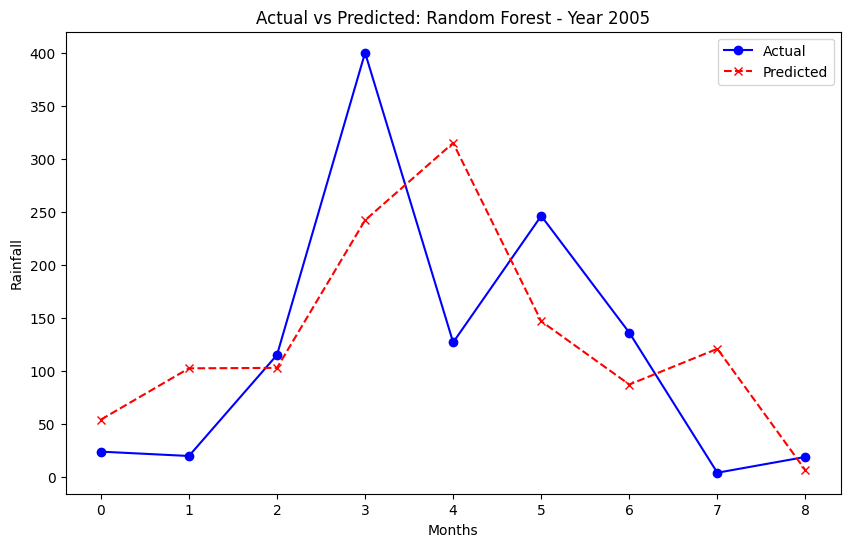
****

****

****

****





**Prediction Observations:**

**Training on complete dataset-**

|  |  |
| --- | --- |
| **Algorithm** | **MAE** |
| Linear Regression | 96.32 |
| SVR | 127.16 |
| Artificial Neural Networks | 92.26 |

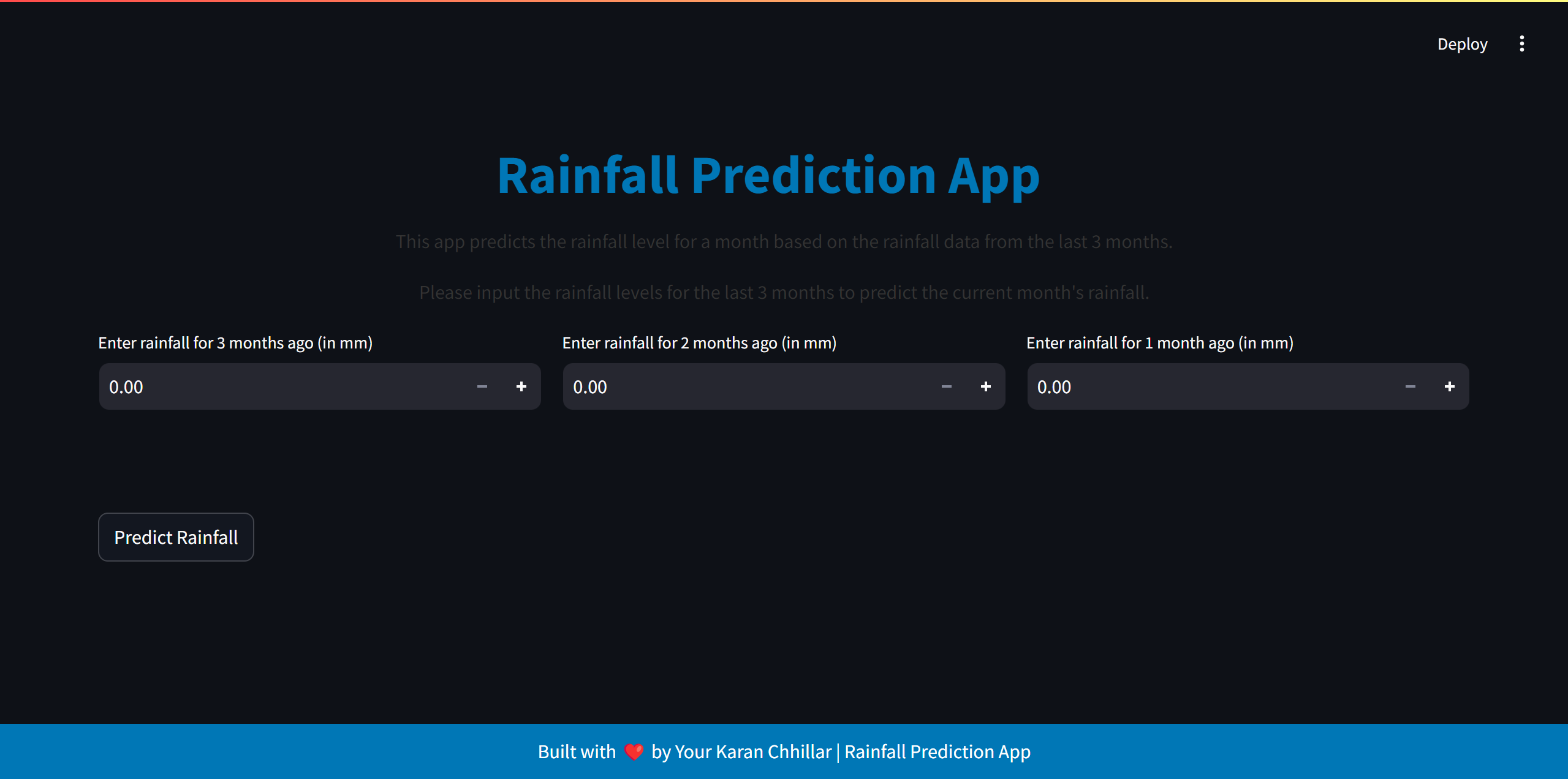
**Technologies:**

• Programming language : Python

• Libraries : numpy, pandas, matplotlib, seaborn, keras, scipy, sklearn

**Web App:**

The web app is built with **Streamlit**, a powerful framework for developing interactive web applications, ensuring an easy-to-use interface and real-time predictions. The model is trained on the Indian Meteorological Department's (IMD) Yearly Gridded Rainfall dataset, offering a comprehensive analysis of the model’s prediction accuracy and statistics for each year.

****

****

**Future Enhancements**

To further enhance the **Rainfall Prediction using Machine Learning** project and improve its predictive capabilities, several key improvements are planned for future development:

**1. Real-Time Rainfall Prediction:**

Integrating real-time data from weather stations or APIs will allow the model to provide up-to-date rainfall forecasts. This will significantly enhance the practical applications of the model for real-world weather forecasting and agricultural planning.

**2. Advanced Model Tuning:**

Further hyperparameter tuning and the application of advanced machine learning techniques, such as ensemble learning and deep learning, will be explored to improve prediction accuracy and model robustness.

**3. Incorporation of Additional Meteorological Factors:**

Incorporating other weather-related parameters such as temperature, humidity, and wind speed into the model will provide a more comprehensive prediction, potentially enhancing the accuracy of rainfall forecasts.

**4. Seasonal and Regional Predictions:**

Developing models specifically tailored to different regions and seasons in India will enable more precise predictions. This approach will address the variability in rainfall patterns across diverse geographical locations and seasonal fluctuations.

**5. Visualization Enhancements:**

Implementing interactive and dynamic visualizations, such as animated heatmaps and time-series graphs, will allow for better interpretation of rainfall trends and predictions. These enhanced visuals can aid users in understanding long-term weather patterns more effectively.

**6. Mobile and Web Interface:**

Developing a user-friendly mobile and web interface will allow users, including farmers and meteorologists, to access the rainfall predictions easily. This will make the project more accessible to a broader audience, enhancing its practical utility.

These enhancements aim to make the project more robust, accurate, and accessible, contributing to the advancement of weather forecasting and climate research.

**Conclusion**

In conclusion, the **Rainfall Prediction using Machine Learning** project represents a significant advancement in leveraging machine learning for weather forecasting. By utilizing historical rainfall data from the Indian Meteorological Department (IMD), the project successfully establishes a foundation for predicting future rainfall patterns. The implemented processes, including data preprocessing, exploratory data analysis, and model selection, have laid the groundwork for building an accurate predictive model.

While the current implementation focuses on model development using machine learning techniques such as Linear Regression, Random Forest, and Neural Networks, the project is poised for further enhancement. Planned improvements, such as incorporating real-time data, refining the model with additional meteorological factors, and improving visualization, will elevate the model’s predictive accuracy and usability.

As the project progresses, these future enhancements will ensure that the rainfall prediction model becomes a valuable tool for weather forecasting, agricultural planning, and disaster management. The ultimate goal is to provide accurate, real-time insights into rainfall trends, contributing to better-informed decisions in climate-sensitive areas.

**References:**

1. Pandas Documentation: https://pandas.pydata.org/docs/
2. NumPy Documentation: https://numpy.org/doc/
3. Scikit-learn Documentation: https://scikit-learn.org/stable/
4. Matplotlib Documentation: https://matplotlib.org/stable/contents.html
5. Indian Meteorological Department (IMD): <https://www.imd.gov.in/>
6. Seaborn Documentation: https://seaborn.pydata.org/