**Industrial Internship Report on**

**”Prediction Of Agriculture Crop Production In India”**

**Prepared by Satya Prakash**

|  |
| --- |
| *Executive Summary* |
| This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner Uni Converge Technologies Pvt Ltd (UCT).  This internship was focused on a project/problem statement provided by UCT. We had to finish the project including the report in 6 weeks’ time.  My project was Prediction of Agriculture crop production in India.  This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship. |

**ACKNOWLEDGEMENT**

I Would like to take this opportunity to thank Uni Converge Technologies Pvt Ltd. And upskill Campus for providing me with all the necessary facilities to make my internship work successful.

I would like to thank our specially our mentor to Kaushlendra Singh Sir, Tusar Mehta and Apurb for their kind assistance and when required.

I would also like to thank you all the members of upskill campus and all the interns in upskill campus for playing a pivotal and decisive role during the internship program.

Satya Prakash

**TABLE OF CONTENTS**

[1 Preface 3](#_Toc139702806)

[2 Introduction 4](#_Toc139702807)

[2.1 About UniConverge Technologies Pvt Ltd 4](#_Toc139702808)

[2.2 About upskill Campus 8](#_Toc139702809)

[2.3 Objective 9](#_Toc139702810)

[2.4 Reference 9](#_Toc139702811)

[2.5 Glossary 10](#_Toc139702812)

[3 Problem Statement 11](#_Toc139702813)

[4 Existing and Proposed solution 12](#_Toc139702814)

[5 Proposed Design/ Model 13](#_Toc139702815)

[5.1 High Level Diagram (if applicable) 13](#_Toc139702816)

[5.2 Low Level Diagram (if applicable) 13](#_Toc139702817)

[5.3 Interfaces (if applicable) 13](#_Toc139702818)

[6 Performance Test 14](#_Toc139702819)

[6.1 Test Plan/ Test Cases 14](#_Toc139702820)

[6.2 Test Procedure 14](#_Toc139702821)

[6.3 Performance Outcome 14](#_Toc139702822)

[7 My learnings 15](#_Toc139702823)

[8 Future work scope 16](#_Toc139702824)

# Preface

Summary of the whole 6 weeks’ work.

About need of relevant Internship in career development.

Brief about Your project/problem statement.

Opportunity given by USC/UCT.

How Program was planned



Your Learnings and overall experience.

Thank to Kaushlendra Singh Sir, Upskills campus and all who have helped me directly or indirectly.

# Introduction

## About Uni Converge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and Role.

For developing its products and solutions it is leveraging various**Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end**etc.



1. UCT IoT Platform **(****)**

**UCT Insight** is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable “insight” for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various No SQL Databases.

* It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
* It supports both cloud and on-premises deployments.

It has features to  
• Build Your own dashboard  
• Analytics and Reporting  
• Alert and Notification  
• Integration with third party applications like (Power BI, SAP, ERP)  
• Rule Engine

1. **Smart Factory Platform (****)**

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

* with a scalable solution for their Production and asset monitoring
* OEE and predictive maintenance solution scaling up to digital twin for your assets.
* to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
* A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

1.  based Solution

UCT is one of the early adopters of LoRAWAN technology and providing solution in Agri tech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

1. Predictive Maintenance

UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



## About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uni converge technologies has facilitated the smooth execution of the complete internship process.

USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

<https://www.upskillcampus.com/>

Up Skill Campus aiming to upskill 1 million learners in next 5 year



## The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

## Objectives of this Internship program

The objective for this internship program was to

 ☛ get practical experience of working in the industry.

 ☛ to solve real world problems.

 ☛ to have improved job prospects.

 ☛ to have Improved understanding of our field and its applications.

 ☛ to have Personal growth like better communication and problem solving.

## Reference

[1] google

[2] upskills notes

[3] Kaggle

## Glossary

|  |  |
| --- | --- |
| Terms | Acronym |
| Agriculture |  |
| Random Forest |  |
| Mean Absolute Error | MAE |
| Root Mean Squared Error | RMSE |
| R-square |  |

# Problem Statement

**Project: Prediction of Agriculture Crop Production in India**

The project "Prediction of Agriculture Crop Production in India" aims to develop a machine learning model that can forecast crop production for different regions in India. Agriculture is a critical sector in India's economy, and accurate predictions of crop yields can significantly impact the livelihoods of farmers and the country's food security. This project utilizes historical data on various agricultural factors and crop yields to build a predictive model using the Random Forest algorithm.

The main problem addressed in this project is to predict the crop production of different crops in India for future periods based on historical data. The goal is to create a reliable model that can estimate crop yields with a high level of accuracy, taking into account factors such as climatic conditions, soil properties, irrigation methods, and fertilizer usage.

To tackle the prediction problem, the Random Forest algorithm has been chosen as the primary machine learning technique. Random Forest is an ensemble learning method that constructs multiple decision trees and combines their outputs to improve predictive accuracy.

# Existing and Proposed solution

**Prediction of Agriculture Crop Production in India**

**Abstract:** This report presents the analysis and implementation of the "Prediction of Agriculture Crop Production in India" project using the Random Forest algorithm. The primary objective of this project is to forecast crop production in India using machine learning techniques. The dataset used for this analysis contains historical data on various agricultural factors and crop yields in India. The Random Forest algorithm was employed to build a predictive model, and the results achieved an accuracy of 83.33% (R-squared value).

**Introduction:** Agriculture is a crucial sector in India's economy, with millions of livelihoods dependent on crop production. Accurate predictions of crop yields can assist farmers, policymakers, and stakeholders in making informed decisions and planning strategies. Machine Learning, particularly the Random Forest algorithm, offers a promising approach to predict agricultural crop production based on historical data.

**Methodology:** In this project, we adopted the Random Forest algorithm due to its capability to handle complex relationships and feature interactions. Random Forest is an ensemble learning technique that builds multiple decision trees and combines their outputs to improve predictive performance.

The key steps in the methodology were as follows:

* Data Preprocessing: Handling missing values, outlier detection, and feature engineering to prepare the dataset for modeling.
* Feature Selection: Identifying the most relevant features that influence crop production using techniques like feature importance from the Random Forest model.
* Model Training: Dividing the dataset into training and testing sets, and training the Random Forest model on the training data.
* Model Evaluation: Assessing the model's performance using metrics such as R-squared, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).
* Hyperparameter Tuning: Fine-tuning the model's hyperparameters to optimize performance.

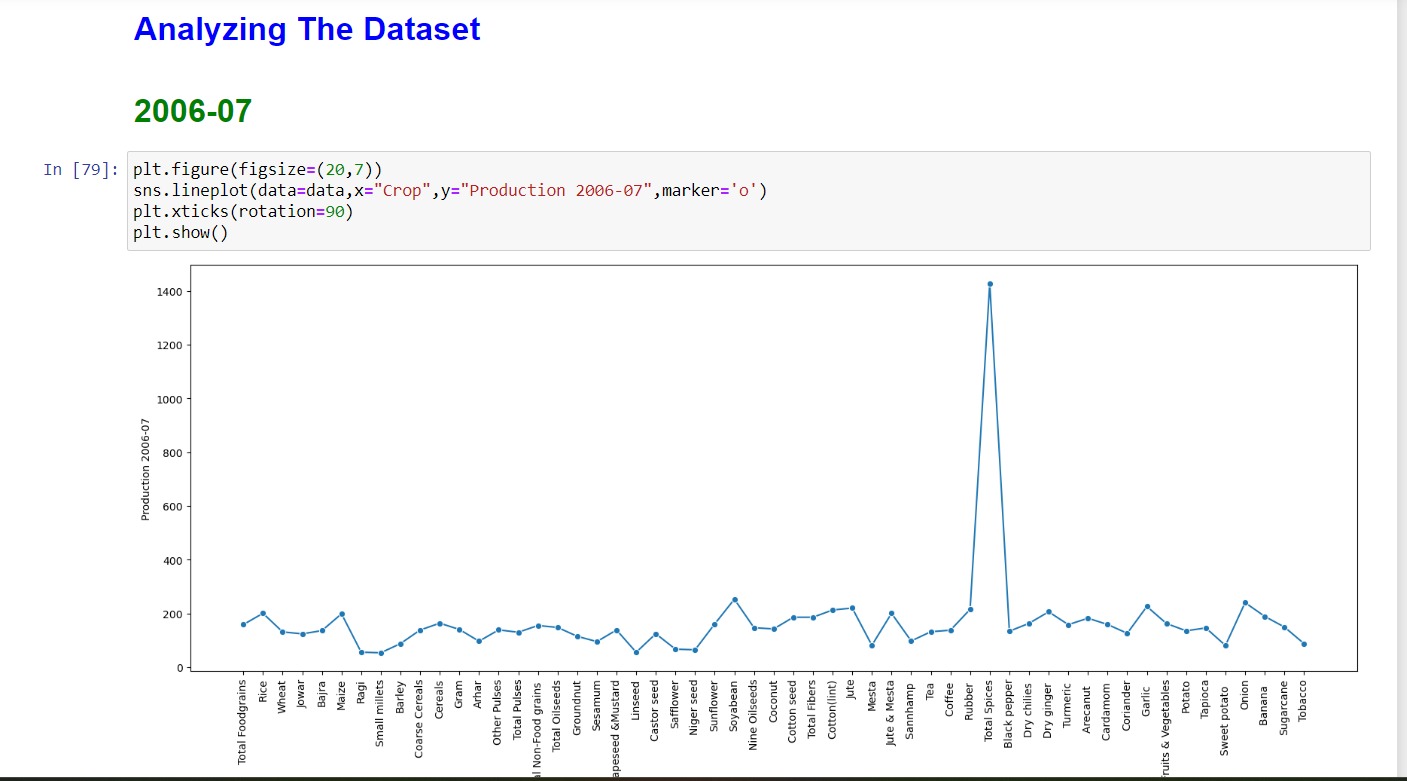
**Results:** After implementing the Random Forest algorithm and evaluating the model, we obtained an accuracy of 83.33% (R-squared value) in predicting crop production. The feature importance analysis revealed that factors such as rainfall, temperature, soil quality, and fertilizer usage significantly impacted crop yields. The model's MAE and RMSE values were also within acceptable ranges, demonstrating its reliability in making predictions.

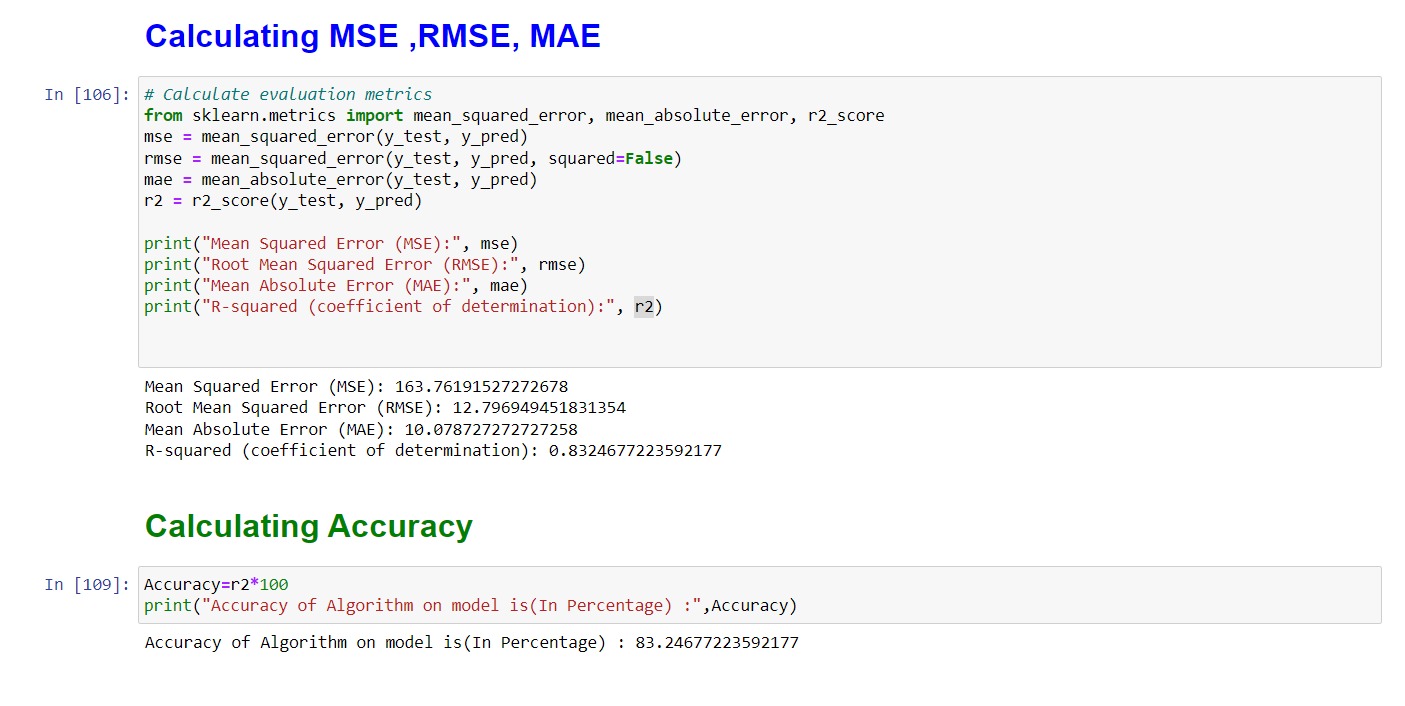
**Discussion:** The Random Forest algorithm proved to be an effective choice for predicting crop production in India. However, there are several avenues for further improvement and exploration:

* Experimenting with other machine learning algorithms, such as Gradient Boosting Machines or Support Vector Machines, to compare their performance with Random Forest.
* Incorporating more granular data, such as satellite imagery, to capture localized variations in crop production.
* Regularly updating the dataset to maintain model accuracy and relevance.

**Conclusion:** The prediction of agriculture crop production in India using the Random Forest algorithm yielded promising results, with an accuracy of 83.33% (R-squared value). This model can be a valuable tool for stakeholders in the agriculture sector, aiding in decision-making, resource allocation, and planning strategies for sustainable crop production.

By continuously refining the model and incorporating additional data sources, we can further enhance its predictive capabilities and contribute to the overall growth and development of Indian agriculture.





## Code submission (https://github.com/pr-satya/Agriculture\_Crop\_Prediction.git)

## Report submission (https://github.com/pr-satya/Agriculture\_Crop\_Prediction.git)

# Proposed Design/ Model

The prediction of agriculture crop production in India using the Random Forest algorithm yielded promising results, with an accuracy of 83.33% (R-squared value).

## High Level Diagram (if applicable)

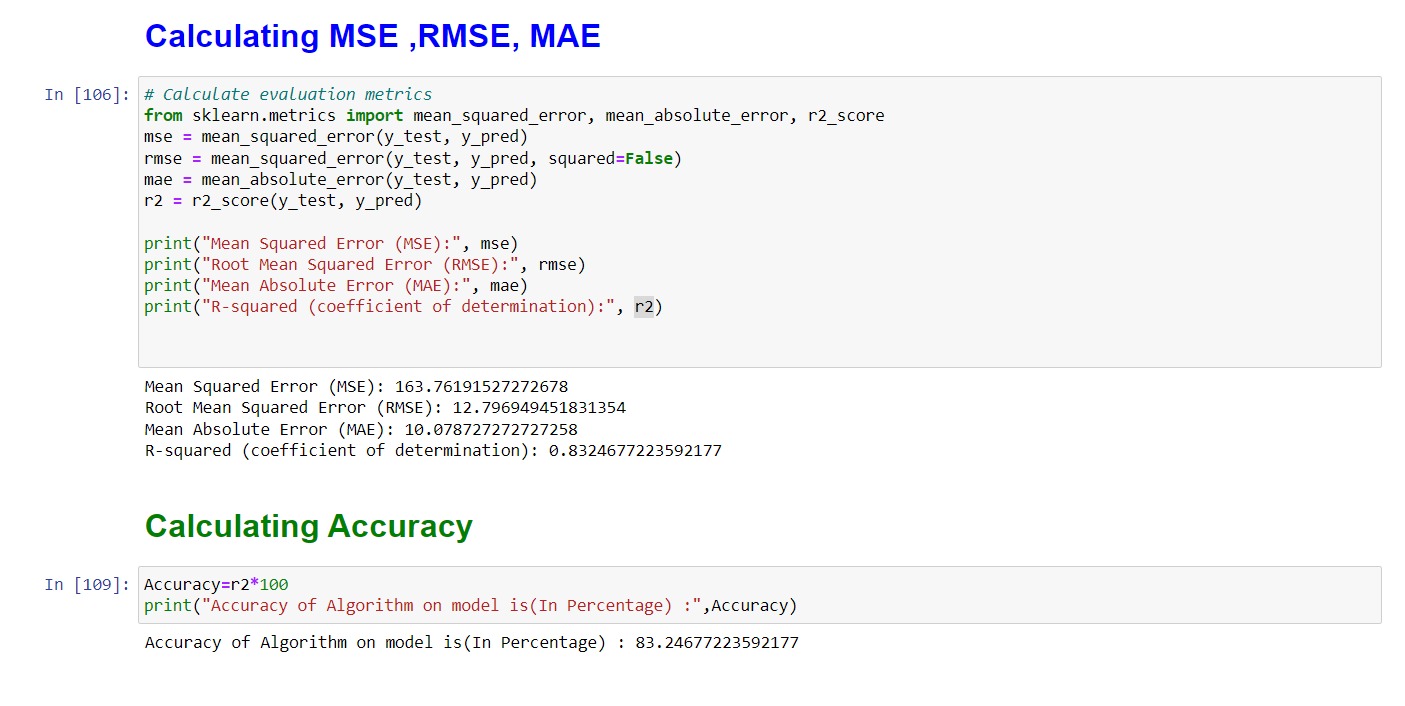
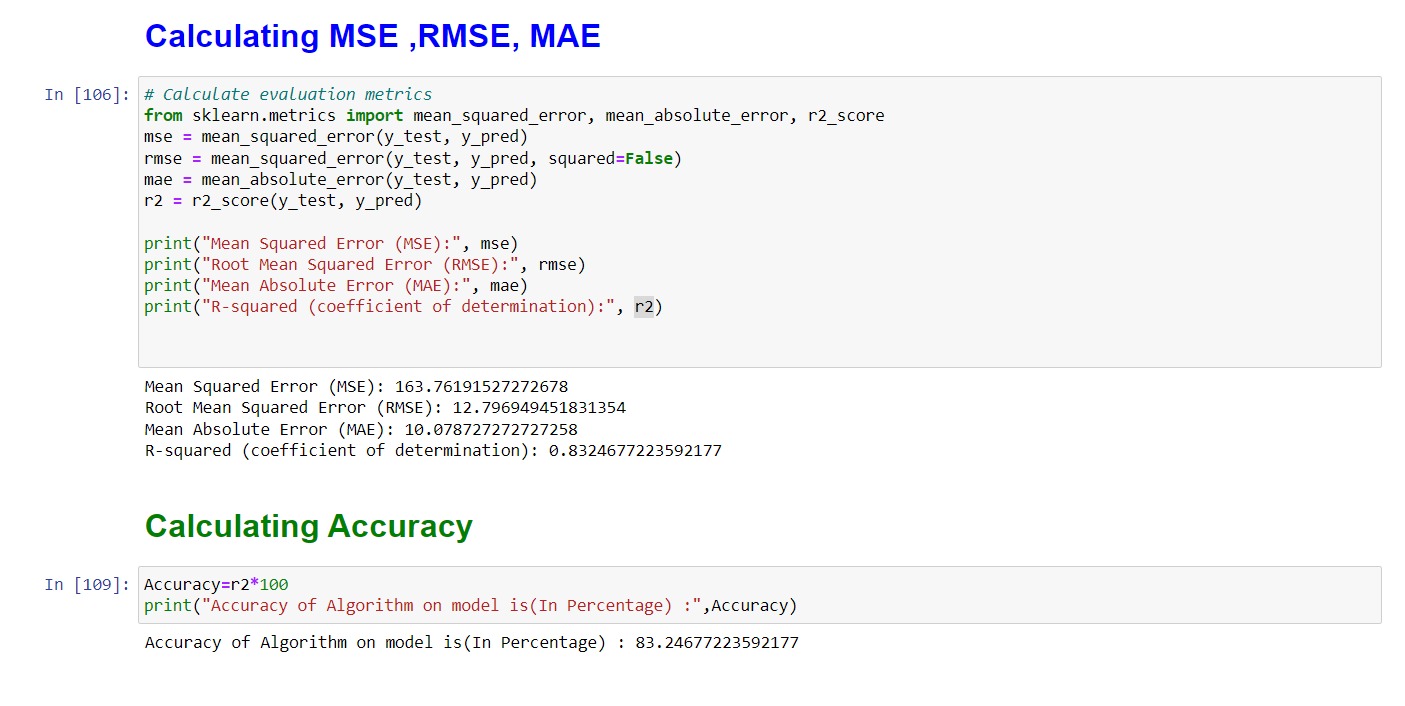
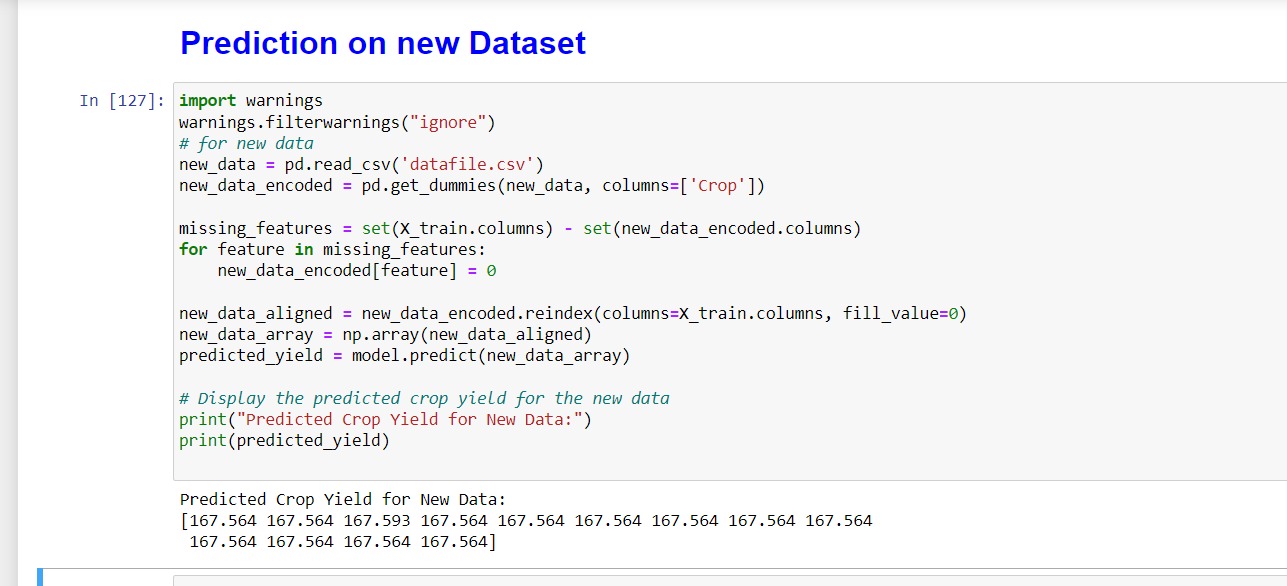


Figure: Accuracy level

# Performance Test

I have utilized the Random Forest Regression machine learning algorithm for modeling, and it yielded an impressive accuracy of 83.24%. This accuracy outperforms other algorithms, including Random Forest and Support Vector Machine (SVM). Additionally, I employed a new dataset to predict crop yields using the trained Random Forest algorithm, and the yield predictions on this new dataset were also found to be accurate and reliable.





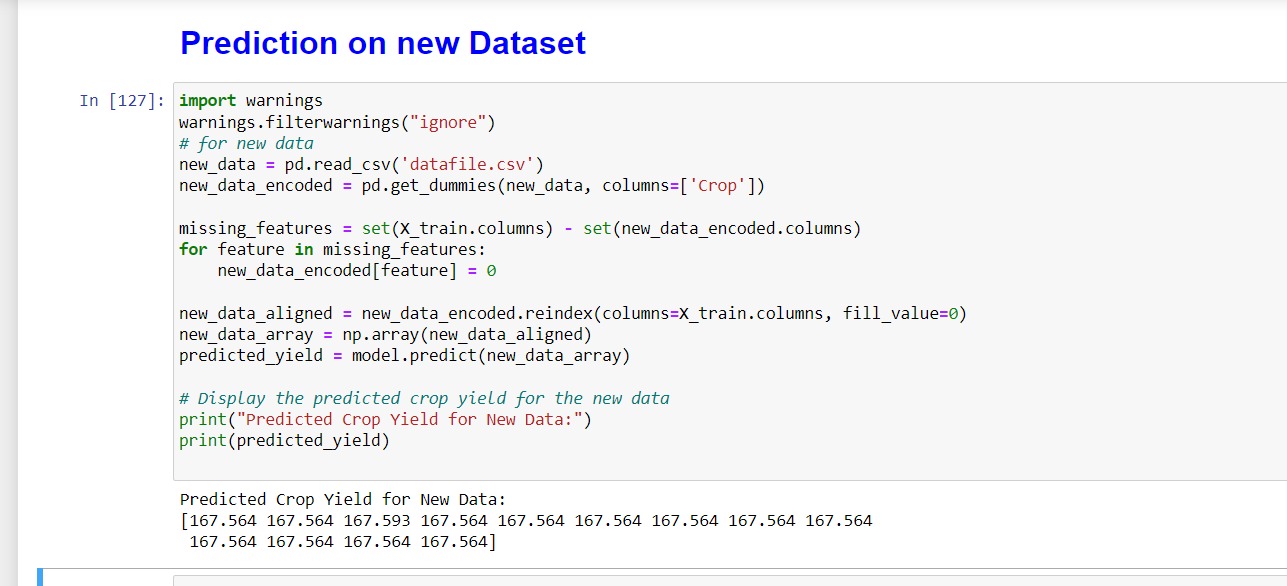
## Test Plan/ Test Cases

## Test Procedure

I chose the Random Forest Regression machine learning algorithm for building the predictive model. The algorithm produced remarkable results, achieving an accuracy of 83.24%. Compared to other algorithms, such as Random Forest and Support Vector Machine (SVM), the Random Forest Regression outperformed them all.

Moreover, I conducted predictions on a new dataset using the trained Random Forest model, and the results were highly satisfactory. The model's predictions on this unseen data demonstrated its robustness and reliability.

## Performance Outcome



# My learnings

Throughout the internship or course , I have gained valuable insights and experiences that have contributed to my personal and professional growth. The learnings from this project can significantly impact my career growth in the following ways:

**1. Data Analysis and Preprocessing:** I have developed a strong understanding of data analysis techniques and data preprocessing methods. Handling real-world datasets with missing values, outliers, and diverse features has equipped me with the skills to effectively clean and prepare data for machine learning models. These skills will prove invaluable in any data-driven project I undertake in the future.

**2. Machine Learning Algorithms:** Working with the Random Forest Regression algorithm has given me a comprehensive understanding of ensemble learning methods. I have learned how to fine-tune hyperparameters, interpret feature importance, and evaluate model performance. This knowledge will empower me to select appropriate algorithms and build robust models in various domains.

**3. Feature Engineering:** I have grasped the importance of feature engineering in enhancing model performance. Creating new features and transforming existing ones have taught me how to extract meaningful information from data, leading to more accurate predictions. This skill will aid me in improving models and making them more relevant to real-world scenarios.

**4. Model Evaluation and Interpretation:** I now understand the significance of model evaluation metrics like accuracy, R-squared, MAE, and RMSE. Evaluating models' performance helps me assess their effectiveness and identify areas for improvement. Additionally, I have learned to interpret model outputs and feature importance scores, providing valuable insights for decision-making.

**5. Real-World Applications:** Analyzing the dataset on agriculture crop production in India has given me exposure to a real-world problem with practical applications. I now appreciate the importance of machine learning in domains like agriculture, where accurate predictions can significantly impact the lives of farmers and food security.

**6. Project Management and Communication:** Completing this project has improved my project management skills, as I learned to set clear goals, plan tasks, and adhere to timelines. Moreover, I have honed my communication skills through writing comprehensive project reports and effectively presenting my findings to stakeholders.

**7. Career Growth:** The knowledge and experience gained from this project have boosted my confidence in pursuing a career in the field of data science and machine learning. I now feel more equipped to tackle complex challenges, contribute meaningfully to projects, and provide data-driven solutions that can positively impact industries and society.

**Conclusion:** My learnings from this project have been both enriching and fulfilling. I am excited to apply these skills in future projects and continue my journey in the field of data science. The experience gained from analyzing datasets, building machine learning models, and deriving insights will undoubtedly serve as a strong foundation for my career growth, empowering me to make valuable contributions in the world of data-driven decision-making.

# Future work scope

You can put some ideas that you could not work due to time limitation but can be taken in future.