**A Minor Project Synopsis**

**on**

**Crop Yield Prediction using ML**

Submitted to Manipal University Jaipur

towards the partial fulfillment for the award of the degree of

**Bachelor of Technology**

**In Information Technology**

By

**Kashish Jaiswal**

219302170

IT-D

**Priya Rani**

219302319

IT-D

Under the Guidance of

**Ms. Shikha Chaudhary**

A close up of a logo

Description automatically generated

**Department of Information Technology**

**School of Information Technology**

**Manipal University Jaipur**

**2022-2023**

**Synopsis**

**1. Introduction**

Agriculture has always been practiced and been a major source of income in India for centuries due to the country’s gifted fertile soils and terrain. About 54% of the population of India is engaged in some form of agricultural practices in the present. Each year, farmers put in tremendous efforts in the hopes of gaining a good amount of yield from their harvests. Hence it becomes important for farmers to know how much yield is expected beforehand.

To help farmers, we are determined to build a Machine Learning Model. In this project we will try to build a model to the best of our knowledge that will make predictions on the amount of yield produced per unit area. Many factors like location of cultivation, rainfall, soil, pesticides usage, type of crop are to be considered that will help in determining and making predictions of yield produced. We will do a comparison analysis of various regression techniques to see which ones give more accuracy.

**2. Motivation**

Farmers work hard day and night to feed their families and people of our country. It includes a lot of physical labor and hefty costs of farming equipments. The motivation behind this project is to help farmers make the right decisions about their farming choices and practices. This project will help the farmers in avoiding unforeseen dangers of low yields which could cause their efforts go in vain. Instead of compensating with the heavy losses caused due to wrong decisions, this model will try to give accurate predictions about yield produced beforehand.

**3. Project Objectives**

1)Crop Yield Prediction: The model will be trained on a pre-existing dataset which contains crucial features related to crop yield prediction like crop-types, cropping seasons, annual rainfall, fertilizer usage etc to predict quantity of crop production per unit area i.e. crop yield

.

2) Analysis of Algorithms: To be able to perform comparison-analysis of different Machine Learning algorithms used in prediction and observe which ones perform better than others.

3) Knowledge Perspective: Through this project we aspire to learn and be able to solve real life problems using emerging technologies like Machine Learning in this project. Being engineers, we aim to automate things and make lives of people in the society, easier.

**4. Methodology/ Planning of work:**

A very general format for any Machine Learning project is as follows:

1. Data Collection
2. Data Cleaning/Pre-processing
3. Data Exploration
4. Data Visualization
5. Implementation of Machine Learning Algorithms and training the model
6. Making Predictions

We are going to follow a very similar approach as above.

* We will be working on a crop yield prediction dataset collected from Indian states. This is a labelled dataset and has been taken from Kaggle website
* Pre-processing or data cleaning will be the next step followed.

Data cleaning is generally done in the following ways:

1. Removing all the NA (missing values) either by deleting the particular entry or by deleting the entire column (if it has significant number of NA values) from the table.
2. Imputation: missing data are filled in to complete data matrix that can be analysed. Missing data is generally filled with most probable/frequent values (or other suitable values).
3. We look for errors in format of a column like unnecessary spaces, comas etc and correct them.

* Next, we form an intuition based on how the data is present. It includes data exploration, visualization in form of various graphs.
* As the given problem is not a classifier problem nor a clustering problem, we will perform Regression analysis.
* Selection of model and implementing Machine Learning Algorithms to calculate precision and accuracy scores.

**5. Facilities required for proposed work:**

Software Requirements

1. Programming Language: Python
2. Libraries used: Pandas, Numpy, Sklearn , Matplotlib etc
3. Dataset: picked from Kaggle website for Crop Yield Prediction
4. Integrated Development Environment (IDE): Google Collab

Hardware Requirements

1. Configurations: 64 bit Intel Core I5 processors, Minimum 8GB RAM machine
2. Operating System: Windows10 and above, Linux, MAC O.S

**Bibliography/References**

1. <https://www.kaggle.com/datasets/crop-yield-in-indian-states-dataset>
2. <https://www.fao.org/home/en>
3. <https://www.analyticsvidhya.com/>
4. <https://www.w3schools.com/datascience/>
5. <https://timesofindia.indiatimes.com/readersblog/shameem-ahmad/the-life-of-an-indian-farmer-36314/>