**INT – 247 Machine Learning Foundation**

A picture containing text

Description automatically generated

Assignment – I

**Agricultural Production Optimization**

Made By

Name - Ankit Kumar

Reg No - 11907132

Section: KM012

**Abstract**

The objective of my project is to create a machine learning project so I decided to do a project on Croups prediction using machine learning which helps to Optimize agricultural production. We start looking for the requirements from the project and start learning from the resources. We looked from some websites like Kaggle, GitHub for seeing some functions need to be included in the program. As, a result I completed the given task in the given amount of time and in while making this project we get to know about various applications that can be performed or made with the help of the functions which we study in python programming language, and we learn how to work with python for machine learning.

**Acknowledgements**

We have taken efforts in this project. However, it would not have been possible without the kind support and help of Mrs. Upinder kaur teaching us machine learning foundation in this semester. I would like to extend our sincere thanks to our mam. We are highly indebted to Upinder kaur mam for their guidance and constant supervision as well as for providing necessary information regarding the project and for their support in completing the project. I would like to express my gratitude towards our parents who always motivates me while doing this project. I would like to thank Lovely Professional University to design and includes such types of projects in every semester.

**Table of contents**

|  |  |  |
| --- | --- | --- |
| sno | Topic | Page.no |
| 1 | INTRODUCTION | 5 |
| 2 | Collecting Dataset | 5 |
| 3 | Pre-processing of data | 6 |
| 4 | Analysing the data | 7 |
| 5 | Building Ml Model | 8 |
| 6 | Modal Evaluation | 9 |
| 7 | References | 12 |

1. **INTRODUCTION**

Predicting the type of crop for a particular area is a very complex and challenging task as it is influenced by several parameters starting from the type of soil and climatic parameters. Again, the crop also depends on the type of method used by the farmers from field to field, so predicting the Crop type’s performance from its parametric point of view is a very critical task. With the increasing population, there is a considerable demand for crops throughout the globe; hence, farmers need to be aware of the type of crop that can be treatable for their geographical location and soil type. This is required to waive the economic pressure on the farmers. Therefore, it is essential to provide accurate, timely-based information based on the climatic parameters and soil type to the farmers, helping them make the best decision for their soil, leading to greater profitability and productivity.

**Steps for Determining the type of croups for a given condition**

1. Collecting Dataset
2. Pre-processing of data
3. Analysing the data
4. Building ML model
5. Evaluating the model
6. **Collecting Dataset**

Data set is collected from Kaggle which have features like

N - ratio of Nitrogen content in soil

P - ratio of Phosphorous content in soil

K - ration of Potassium content in soil

temperature - temperature in degree Celsius

humidity - relative humidity in %

ph - ph value of the soil

rainfall - rainfall in mm

A black screen with white text

Description automatically generated with low confidence

1. **Pre-processing of data**

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So for this, we use data preprocessing task.

It involves below steps:

* Getting the dataset
* Importing libraries
* Importing datasets
* Finding Missing Data
* Encoding Categorical Data
* Splitting dataset into training and test set
* Feature scaling

1. **Analysing the data**

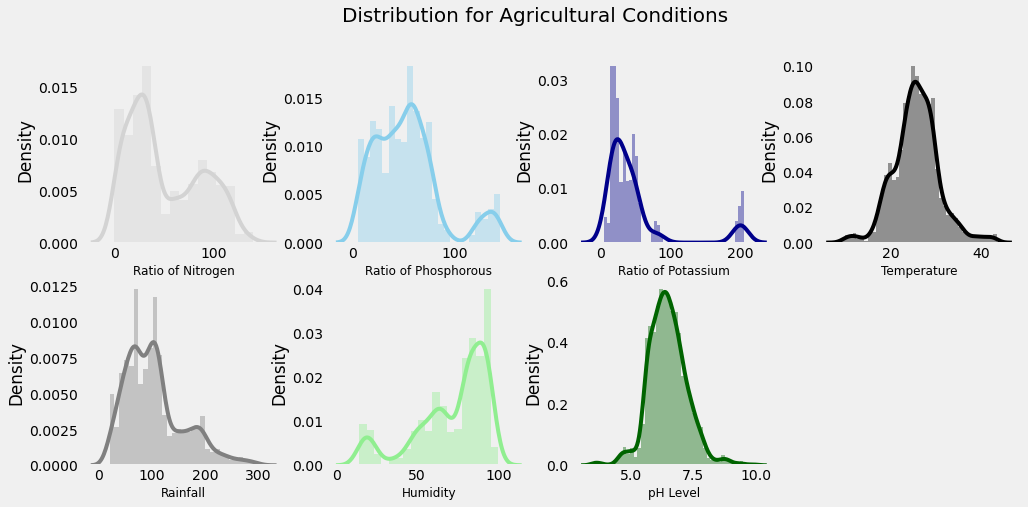
There are several **types of Data Analysis** techniques that exist based on business and technology. However, the major Data Analysis methods are:

* Text Analysis
* Statistical Analysis
* Diagnostic Analysis
* Predictive Analysis
* Prescriptive Analysis
  1. **Descriptive Statistics**

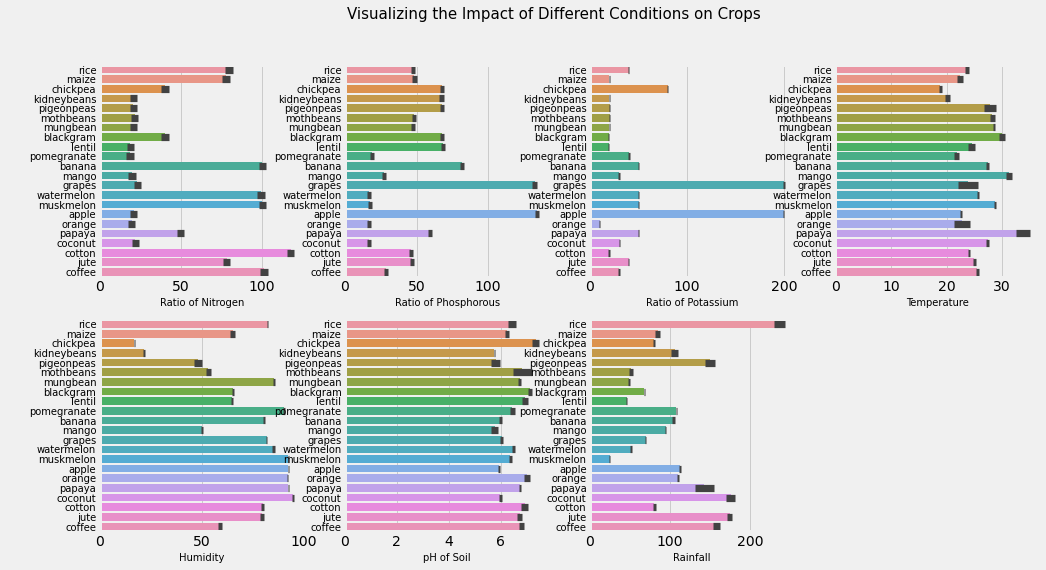
Text

Description automatically generated

* 1. **distribution of Agricultural Conditions**



* 1. **Impact of Different Conditions on Crops**



1. **Building Ml Model**

I have used Logistic Regression for Building The ML Model

Logestic Regression: Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on.

Text

Description automatically generated

1. **Modal Evaluation**

Model Evaluation is the process through which we quantify the quality of a system’s predictions. To do this, we measure the newly trained model performance on a new and independent dataset. This model will compare labeled data with it’s own predictions.

**Model evaluation performance metrics teach us:**

* How well our model is performing
* Is our model accurate enough to put into production
* Will a larger training set improve my model’s performance?
* Is my model under-fitting or over-fitting?

## **Metrics for classification models**

The following metrics are reported when evaluating classification models:

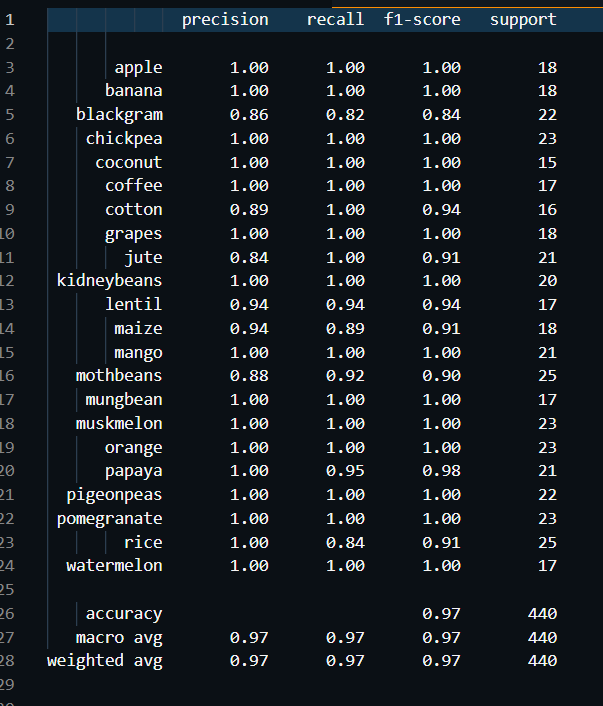
* **Accuracy** measures the proportion of true results to total cases. Aim for a high accuracy rate.  
  accuracy = correct predictions / total data points
* **Log loss** is a single score that represents the advantage of the classifier over a random prediction. The log loss measures the uncertainty of your model by comparing the probabilities of it’s outputs to the known values (ground truth).. we want to minimize log loss for the model as a whole.
* **Precision** is the proportion of true results over all positive results.
* **Recall** is the fraction of all correct results returned by the model.
* **F1-score** is the weighted average of precision and recall between 0 and 1, where the ideal F-score value is 1.
* **AUC** measures the area under the curve plotted with true positives on the y axis and false positives on the x axis. This metric is useful because it provides a single number that lets you compare models of different types.
* **Confusion Matrix** the correlation between the label and the model’s classification. One axis of a confusion matrix is the label that the model predicted, and the other axis is the actual label.

A screenshot of a computer

Description automatically generated with medium confidence

Background pattern

Description automatically generated



1. **References**

* Kaggle.com
* Youtube.com
* Udemy.com
* Google.com