**IMPLEMENTATION**

**MODULES:**

* Data Collection
* Dataset
* Data Preparation
* Model Selection
* Analyze and Prediction
* Accuracy on test set
* Saving the Trained Model

**MODULES DESCSRIPTION:**

**Data Collection:**

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform.

There are several techniques to collect the data, like web scraping, manual interventions and etc.

The dataset used in this crop recommendation in india taken from some other source

**Dataset:**

The dataset consists of 821  individual data. There are 14 columns in the dataset, which are described below.

1. ***States***: total number of states in india
2. ***Rainfall***: rainfall in mm
3. ***Ground Water***: Total ground water level
4. ***Temperature***: temperature in degree Celsius
5. ***Soil type***: Number of soil types
6. ***Season***: Which season is suitable for crops
7. ***Crops***: Types of crops
8. ***Fertilisers required*** : Types of Fertilisers required
9. ***Cost of cultivation***: Total cost for cultivation
10. ***Expected revenues***: Total expected revenues
11. ***Quantity of seeds per hectare***: seeds for quantity per hectare
12. ***Duration of cultivation***: number of day for duration of cultivation
13. ***Demand of crop***: demand of crop (High, low)
14. ***Crops for mixed cropping***: which crop can mixed for cropping

**Data Preparation:**

Wrangle data and prepare it for training. Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.)

Randomize data, which erases the effects of the particular order in which we collected and/or otherwise prepared our data

Visualize data to help detect relevant relationships between variables or class imbalances (bias alert!), or perform other exploratory analysis

Split into training and evaluation sets

**Model Selection:**

A decision tree is a flowchart-like tree structure where an internal node represents feature(or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in recursively manner call recursive partitioning. This flowchart-like structure helps you in decision making. It's visualization like a flowchart diagram which easily mimics the human level thinking. That is why decision trees are easy to understand and interpret.

Decision Tree is a white box type of ML algorithm. It shares internal decision-making logic, which is not available in the black box type of algorithms such as Neural Network. Its training time is faster compared to the neural network algorithm. The time complexity of decision trees is a function of the number of records and number of attributes in the given data. The decision tree is a distribution-free or non-parametric method, which does not depend upon probability distribution assumptions. Decision trees can handle high dimensional data with good accuracy.

The decision rules are generally in form of if-then-else statements. The deeper the tree, the more complex the rules and fitter the model.

Before we dive deep, let's get familiar with some of the terminologies:

* Instances: Refer to the vector of features or attributes that define the input space
* Attribute: A quantity describing an instance
* Concept: The function that maps input to output
* Target Concept: The function that we are trying to find, i.e., the actual answer
* Hypothesis Class: Set of all the possible functions
* Sample: A set of inputs paired with a label, which is the correct output (also known as the Training Set)
* Candidate Concept: A concept which we think is the target concept
* Testing Set: Similar to the training set and is used to test the candidate concept and determine its performance

**Analyze and Prediction:**

In the actual dataset, we chose only 7 features :

1. ***States***: total number of states in india
2. ***Rainfall***: rainfall in mm
3. ***Ground Water***: Total ground water level
4. ***Temperature***: temperature in degree Celsius
5. ***Soil type***: Number of soil types
6. ***Season***: Which season is suitable for crops
7. ***Crops***: Types of crops

**Accuracy on test set:**

We got a accuracy of 90.7% on test set.

**Saving the Trained Model:**

Once you’re confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or . pkl file using a library like pickle .

Make sure you have pickle  installed in your environment.

Next, let’s import the module and dump the model into.pkl file