`

DATA MINING and MACHINE LEARNING PROJECT

BIN 203

PREDICTIVE ANALYSIS OF DIGITAL AGRICULTURE

Delphina Sherine,Jeevitha V,Vijayalakshmi A B

(122013009,122013019,122013053)

II year,BTech Bioinformatics

`

PROBLEM DESCRIPTION

Cultivation of a crop in a location without the proper knowledge of the climatic conditions or without analysis of whether that climatic conditions support the growth of that crop or not may not give proper results and may lead to loss of time and money.Digital agriculture is a method of prior prediction to determine whether a particular crop can grow in a specified location or not depending on the temperature, rainfall and soil conditions of that area .This can be useful to check whether the cultivation of that crop is really viable ,profitable to the farmers in that particular climatic condition in advance. This hence prevents the loss of money and save time, in case of bad yield.
Data mining plays a major role here. It is used to find the relationship between the various factors(attributes) affecting crop growth which hence helps in predicting whether the growth of that particular crop is apt for that location and hence gives good yield, profit to the farmer.

INTRODUCTION TO DATASET

no	CROP	Favourable areas of growth in India	Temperature(in Celsius)	Rainfall(cm)	Type of soil required for its optimum growth	Class Label
1	WATERMELON	Uttar Pradesh, Himachal Pradesh, Rajasthan, Orissa, Gujarat, Punjab, Haryana, Assam, West Bengal, Karnataka, Orissa, Andhra Pradesh, Maharashtra, and Tamil Nadu	22-28	40-60	sandy,light soil	yes
2	MUSKMELON	Punjab, Tamil Nadu, Uttar Pradesh, Maharashtra, and Andhra Pradesh.	25°C - 35°C	Liberal	sandy,loam	yes
3	CUCUMBER	Cucumber grows well in most of all states that has good sunlight,adequatedrainage,rich fertile soil	21c	80	loose sandy loam soil	yes
4	BITTERGOURD	Andhra Pradesh, Odisha, Bihar, Chhattisgarh, Madhya Pradesh, Assam	24 to 27c	Liberal	sandy loam	yes
5	WHEAT	Uttar Pradesh, Punjab, Haryana, MP	10° to 15°C in	75	flat alluvial soil	Yes

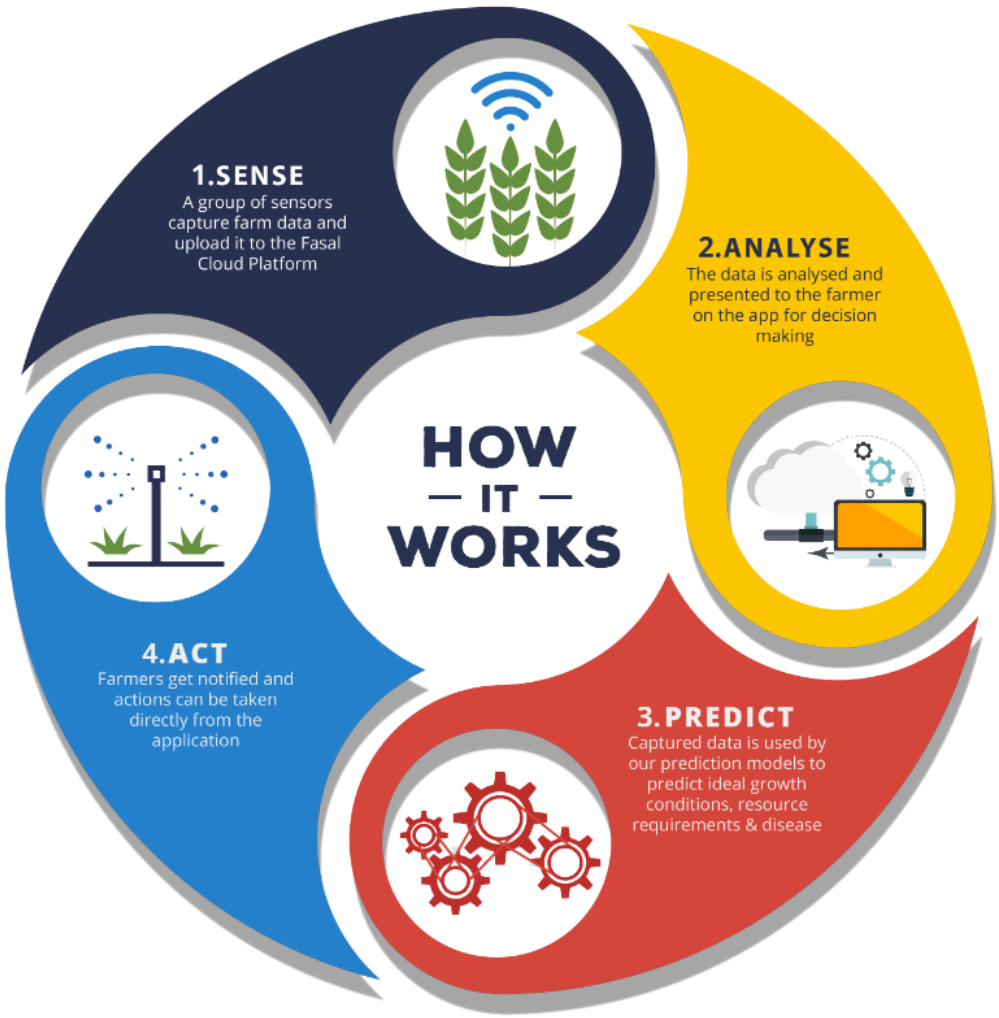
`

6	BARLEY	Uttar Pradesh, Rajasthan, and Madhya Pradesh	18-25	40-45	Sandy Loam, Loam and Medium & Heavy Black Soils.	yes
7	MUSTARD	Rajasthan and Uttar Pradesh Madhya Pradesh, Haryana, Gujarat, West Bengal and Assam.	10 to 20 c	65-100	light to heavy loamy soils	yes
8	PEAS	Himachal Pradesh, Madhya Pradesh, Rajasthan, Maharashtra, Punjab, Haryana, Karnataka and Bihar.	12-20 c	50	Compost soil	yes
9	SESAME	Maharashtra, Rajasthan, West Bengal, Andhra Pradesh, Gujarat, Tamil Nadu, Madhya Pradesh, and Telangana.	25	50-65	well-drained, fertile soils of medium texture and neutral pH.	yes
10	RICE	West Bengal,uttarPradesh,AndhraPradesh,Punjab	more than 25c	More than 100	alluvial soil, mixed soil, clayey soil, loamy soil	yes
11	MAIZE	Karnataka, Andhra Pradesh, Tamil Nadu, Rajasthan, Maharashtra, Bihar, Uttar Pradesh	between 21 and 27	50 to 100	alluvial, red loams	yes
12	COTTON	Punjab, Haryana and Rajasthan, Andhra Pradesh, Tamil Nadu and Karnataka, Gujarat, Maharashtra	21 and 30	50 to 100	Black soil,alluvialsoil,red soil	yes
13	SOYABEAN	Madhya Pradesh, Maharashtra and Rajasthan	20-30	40	Clay loam,alluvial	yes
14	GROUNDNUT	Gujarat,AndhraPradesh,Tamilnadu,Karnataka	30	50	Sandy loam soil	yes

`

Motivation and Introduction:

Digital agriculture is the use of **digital** technology to integrate agricultural production from the paddock to the consumer. These technologies can provide the **agricultural** industry with tools and information to make more informed decisions and improve productivity.



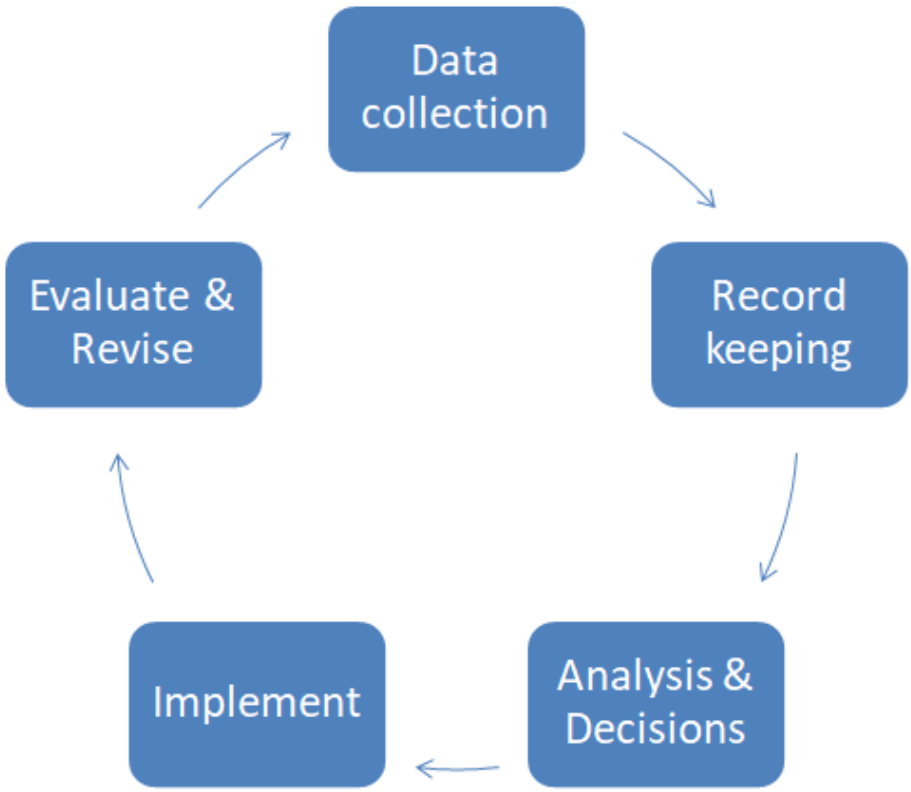
This technology allows for improved crop production by understanding soil health. It allows farmers to use fewer pesticides on their crops. Soil and weather monitoring reduces water waste. **Digital agriculture** ideally leads to economic growth by allowing farmers to get the most production out of their land.

`

Digital techniques involved in this technology:

- **Cloud computing** – enables seamless data storage and real-time reporting across the value chain.
- **Analytics** –it turns vast amount of data into actionable information and knowledge.
- **Internet of Things** – stitch together diverse sources of information.
- **Breeding Informatics**–accelerates R & D for genetic gain.
- **Digital services** – enable targeted provision of farmer preferred products and services.
- **GIS & UAVs –** provides spatial and temporal dimension to information.

Simple flow chart:

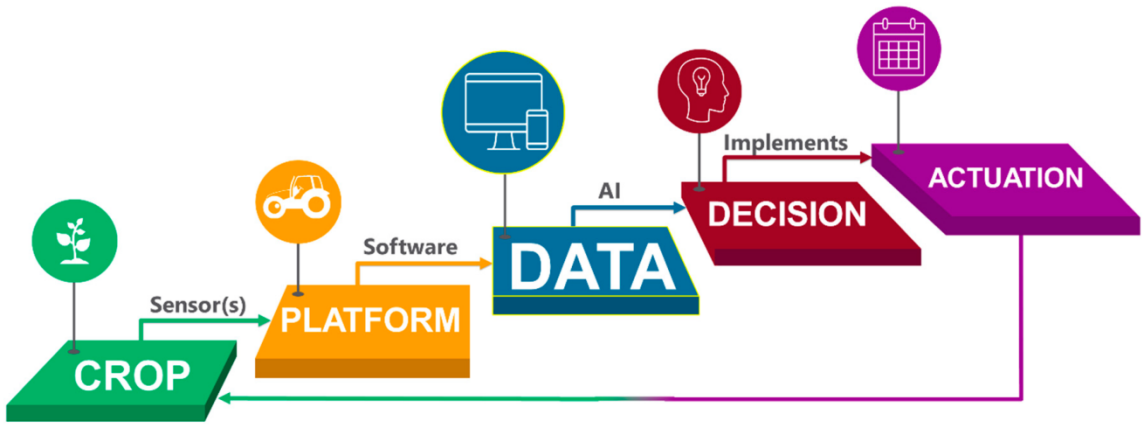


`

Objectives:

- To increase the accuracy and efficiency of agricultural input applications.
- To conduct lab and field experiments and researches on the economic and environmental negative impacts of agricultural chemical misapplications.
- To conduct researches on the economic and environmental feasibility of the precision farming technology that would promote this technology to be accepted and adopted by farmers and environmentalists .
- To localize this technology through collaborative research efforts.
- To deliver this technology to a wide segment of local researchers, environmentalists, farmers and agro-economists.

METHODS



Digital agriculture is a method of prior prediction to determine whether a particular crop can grow in a specified location or not depending on the temperature, rainfall and soil conditions of that area .This can be useful to check whether the cultivation of that crop is really viable ,profitable to the farmers in that particular climatic condition in advance.This can be achieved by data mining and machine learning techniques.
Machine learning is defined by three parameters:

Task(T),Performance(P),Experience(E).The T,P,E for this machine learning problem is:
Task-T-To find a whether a particular crop can grow properly in that climatic condition or not
Performance-P-the ability to correctly classify the growth of a crop in a particular region.

`

Experience-E-database of crop growth in different climatic and soil conditions

Primarily,sensors are used in the agricultural field area that enables seamless data storage and real-time reporting across the value chain by cloud computing. Data here,refers to the factors like weather conditions and soil conditions that affect crop growth.These data are then analysed by machine learning techniques to understand the growth of the crop and to know whether there will be proper harvest or not.

Machine learning is hence performed with input as the real-time reporting data by the help of sensors.This data is fed to a machine learning algorithm like naive-bayes or decision tree algorithm to generate a machine learning model by using WEKA.Evaluation is done by calculating its accuracy,precision and F1-measure.The TP,FP,FN.FP values required for the calculation of the accuracy measures can be evaluated and the confusion matrix is generated using these models developed by the algorithms.

Accuracy	(TP + TN) / (TP + TN + FP + FN)	The percentage of predictions that are correct
Precision	TP / (TP + FP)	The percentage of positive predictions that are correct
Sensitivity (Recall)	TP / (TP + FN)	The percentage of positive cases that were predicted as positive
Specificity	TN / (TN + FP)	The percentage of negative cases that were predicted as negative

$$F_1=2pr/(p+r)$$

If the model so generated is found to be reasonable,accurate and generic without overfitting and outliers,the model can be hence used for further decision making purposes.This performance of the machine improves with experience

Hence machine learning is used to pre-determine whether there will be proper growth of crop in a given set of climatic and soil conditions.This can be used for decision making by the farmers,agricultural industries and related fields of agriculture to take precise decisions in a prior stage.These decisions when implemented to real life situations can increase the performance of agricultural industry as a whole and hence improve food production.

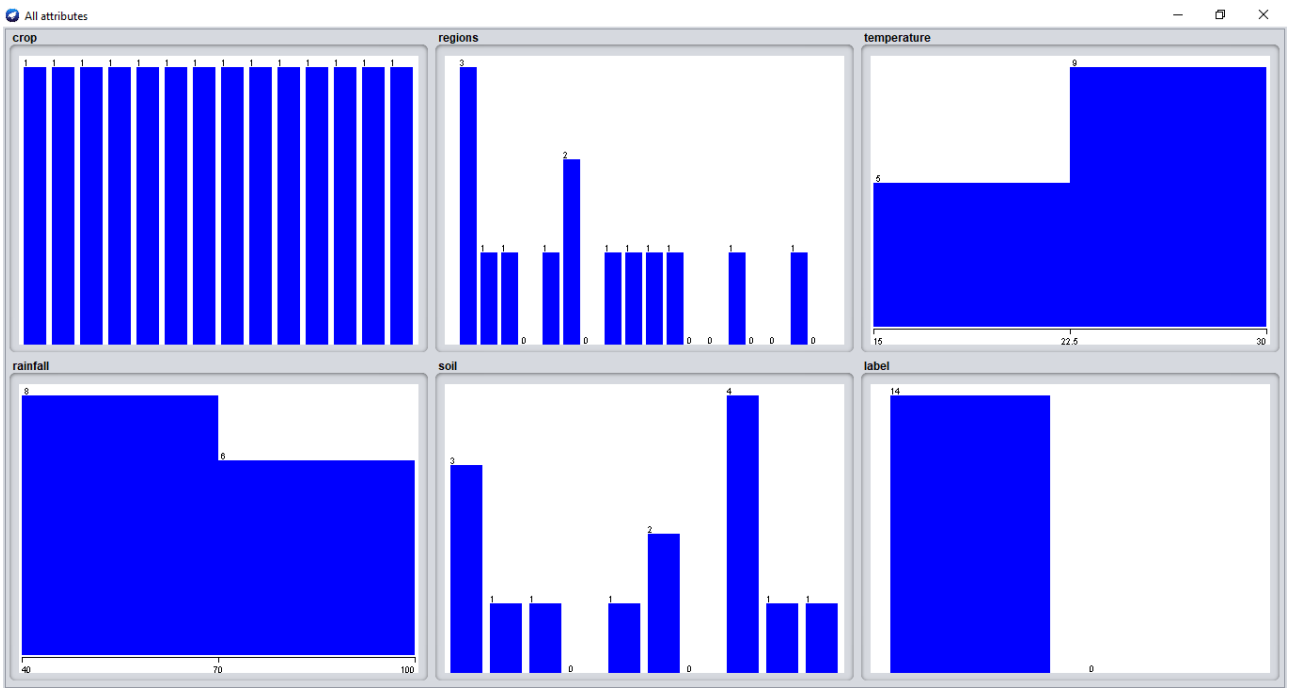
`

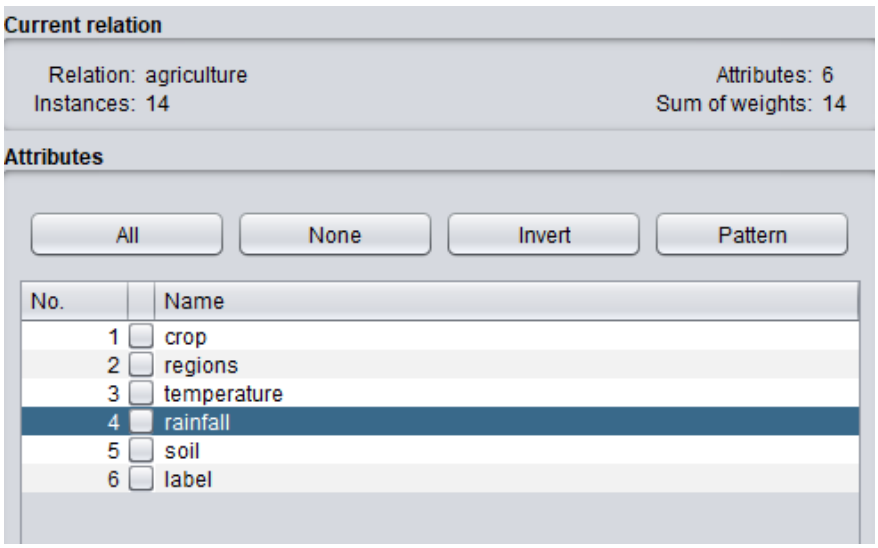
MODEL:

WEKA is a collection of machine learning algorithms for machine learning tasks.It contains tools for data pre-processing, classification,regression, clustering ,association rules and visualisation. The machine learning algorithms-Decision tree and naive-bayes is implemented using **WEKA** for our dataset
The model can be evaluated by predicting the accuracy and F1 measures.The graphs and mining model of our dataset is as shown below.

HISTOGRAMS-VISUALISATION

Histogram for attribute distribution for a single selected attribute can be visualized





`

IMPLEMENTATION OF NAÏVE-BAYES ALGORITHMS

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     agriculture
Instances:    14
Attributes:   6
              crop
              regions
              temperature
              rainfall
              soil
              label
Test mode:    10-fold cross-validation
```

Naive Bayes Classifier

Attribute	Class yes (0.94)	no (0.06)
crop		
rice	2.0	1.0
maize	2.0	1.0
cotton	2.0	1.0
soyabean	2.0	1.0
groundnut	2.0	1.0
watermelon	2.0	1.0
muskmelon	2.0	1.0
cucumber	2.0	1.0
bittergourd	2.0	1.0
wheat	2.0	1.0
barley	2.0	1.0
mustard	2.0	1.0
peas	2.0	1.0
sesame	2.0	1.0
[total]	28.0	14.0
regions		
UP	4.0	1.0
HP	2.0	1.0
Rajasthan	2.0	1.0
Orissa	1.0	1.0
Gujarat	2.0	1.0
Punjab	3.0	1.0
Assam	1.0	1.0
WB	2.0	1.0
Karnataka	2.0	1.0
AP	2.0	1.0
Maharashtra	2.0	1.0
Tamil	1.0	1.0
Nadu	1.0	1.0
all	2.0	1.0
Bihar	1.0	1.0
Chhattisgarh	1.0	1.0
MP	2.0	1.0
Haryana	1.0	1.0
[total]	32.0	18.0
temperature		
mean	23.4375	0
std. dev.	4.0556	0.3125
weight sum	14	0
precision	1.875	1.875

`

rainfall		
mean	63.2143	0
std. dev.	17.1763	1.25
weight sum	14	0
precision	7.5	7.5
soil		
alluvial	4.0	1.0
clay	2.0	1.0
loam	2.0	1.0
mixed	1.0	1.0
black	2.0	1.0
sand	3.0	1.0
light	1.0	1.0
sandyloam	5.0	1.0
compost	2.0	1.0
fertile	2.0	1.0
[total]	24.0	10.0

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          14               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0      %
Root relative squared error              0      %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Cla
                 1.000    ?        1.000      1.000    1.000      ?        ?         1.000     yes
                 ?        0.000    ?          ?        ?          ?        ?         ?         no
Weighted Avg.    1.000    ?        1.000      1.000    1.000      ?        ?         1.000

=== Confusion Matrix ===

  a  b   <-- classified as
 14  0 |  a = yes
  0  0 |  b = no
```

`

IMPLEMENTATION OF DECISION TREE ALGORITHM:

```
Scheme:       weka.classifiers.rules.DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1
Relation:     agriculture
Instances:    14
Attributes:   6
              crop
              regions
              temperature
              rainfall
              soil
              label
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

Decision Table:

Number of training instances: 14
Number of Rules : 1
Non matches covered by Majority class.
        Best first.
        Start set: no attributes
        Search direction: forward
        Stale search after 5 node expansions
        Total number of subsets evaluated: 19
        Merit of best subset found:  100
Evaluation (for feature selection): CV (leave one out)
Feature set: 6

Time taken to build model: 0.02 seconds
```

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          14               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0.0746
Root mean squared error                  0.0746
Relative absolute error                107.4661 %
Root relative squared error            107.4755 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Clas
                 1.000    ?        1.000      1.000    1.000      ?        ?         1.000     yes
                 ?        0.000    ?          ?        ?          ?        ?         ?         no
Weighted Avg.    1.000    ?        1.000      1.000    1.000      ?        ?         1.000

=== Confusion Matrix ===

  a  b   <-- classified as
 14  0 |  a = yes
  0  0 |  b = no
```

`

Results:

Models in machine learning was developed by using WEKA.Histogram for attribute distribution for a single selected attribute was visulaized by WEKA.The dataset was classified using machine learning models developed from algorithms like decision tree and naive-bayes.

Reason for the selection of digital agriculture:

Digital agriculture is the use of new and advanced technologies, integrated into one system, to enable farmers and other stakeholders within the agriculture value chain to improve food production.

Advantages of digital agriculture:

- increase yield
- targeted irrigation
- giving the power of information
- sharing of information
- improving land management
- improving fish farming
- improving supply chains
- monitoring waste production

DISCUSSION AND CONCLUSION:

- The identification of dataset and was analysed
- The machine learning classification algorithm-decision tree and naive-bayes to be used was discussed
- The analysis of dataset was done using charts and graphs.the positive and negatively skewed graphs were considered as the best method to explain the distribution of rainfall and temperature against the crops.
- The feature selection was performed using the filter approach method.Since it reduces the dimensionality of data.
- The outcome attribute is to be predicted for the test dataset by the machine learning model
- The features were also created using feature construction method.
- the naïve bayes and decision tree algorithm were used in our project using the weka model.

`

REFERENCES :

For Dataset:

https://toolbox.google.com/datasetsearch

www.kaggle.com

For dataset information:

www.google.com

www.apnikheti.com

www.kisansuvidha.com