Crop data analytics using image and non-image features

Team members	:	Asish Varanasi Dheeraj Raghavendra Purnachand Jaddu
Mentors	:	Dr. Avinash Sharma Dr. P. Krishna Reddy

Motivation

- Agriculture plays a vital role in economy for developing countries like India.
- Rapid technological advancement - but not much focus on agricultural domain.
- Several newly emerging problems in agriculture.
- In 2012, the NCRB of India reported 13,754 farmer suicides.
- Hence, a need for guidance to the farmers.
- Esagu is one such platform.

Key pain points

Recommendation of Diagnostics:

- Accuracy
- Timely advice
- Minimum required input
- Ease of communication
- Cost effectiveness
- Feedback system

Insights mining

Periodic behaviour of diseases with respect to :

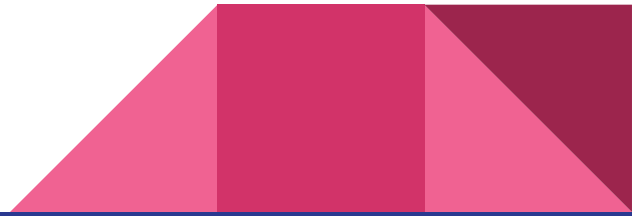
- Season
- Crop
- Pesticides/Fertilizers
- Location
- Pests

Visualization

- Heatmaps showing the variation in the intensity of factors w.r.t:
  - ➢Time
  - ➢Location
- Plots like Bar charts, Pie diagrams etc.
- A dashboard for interactive visualization of the data.

Data

- Input data primarily contains non-image features.
- Weather details, soil conditions and farm report constitute the major part.
- Pre-processing involving data cleaning, integration and normalization.
- Finally, analyzation after data transformation.



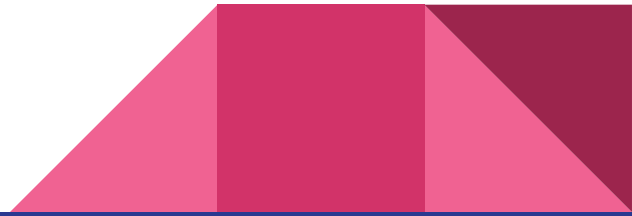
Representation

PGM (Probabilistic Graphical Model)

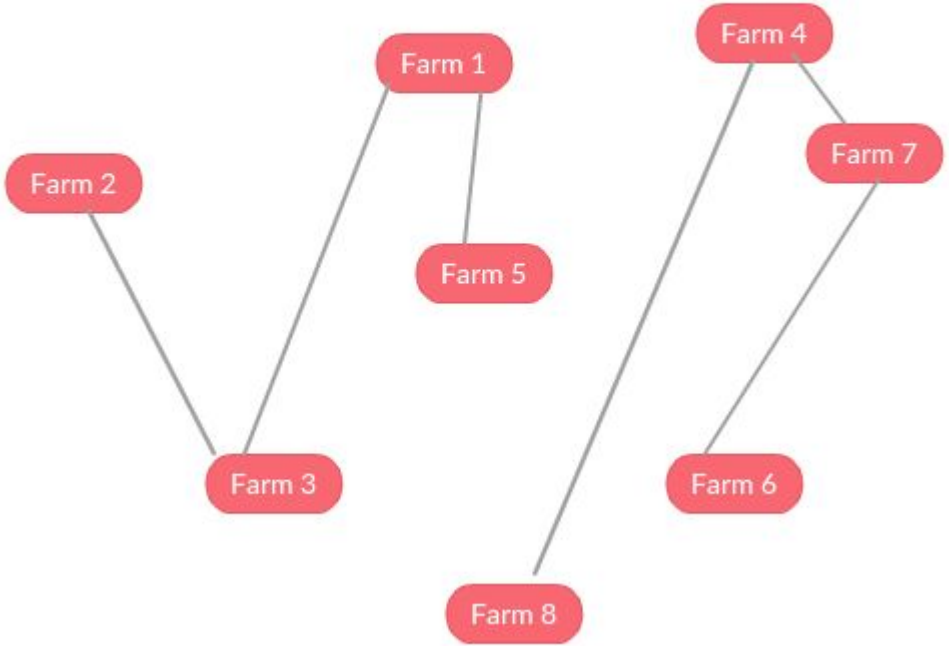
- Farms as vertices and edges indicating the dependency on various factors.

Graph signalling

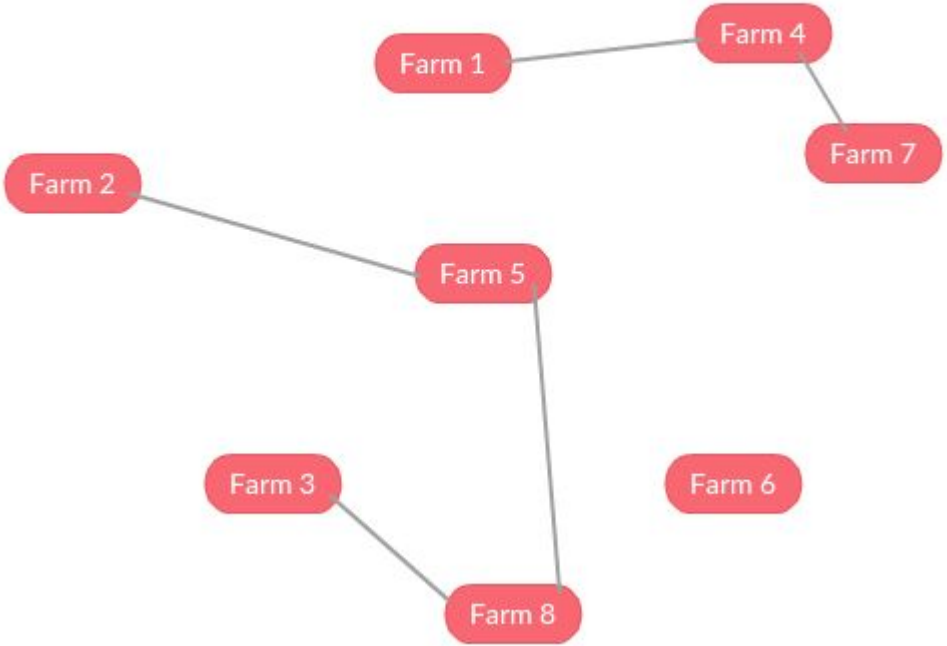
- A plot of the condition of each farm with respect to time/season.



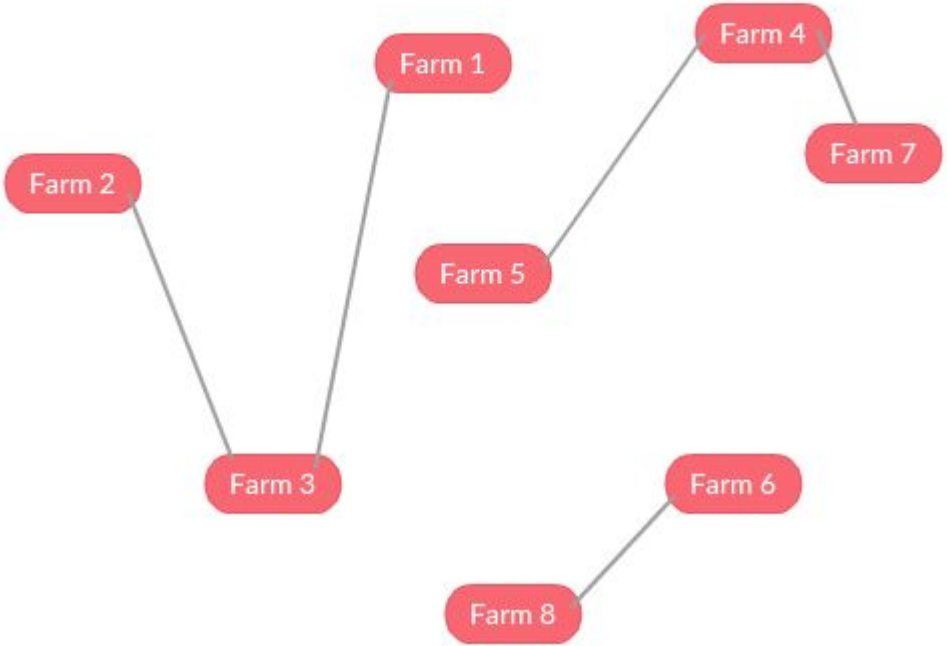
Graph based on neighbourhood



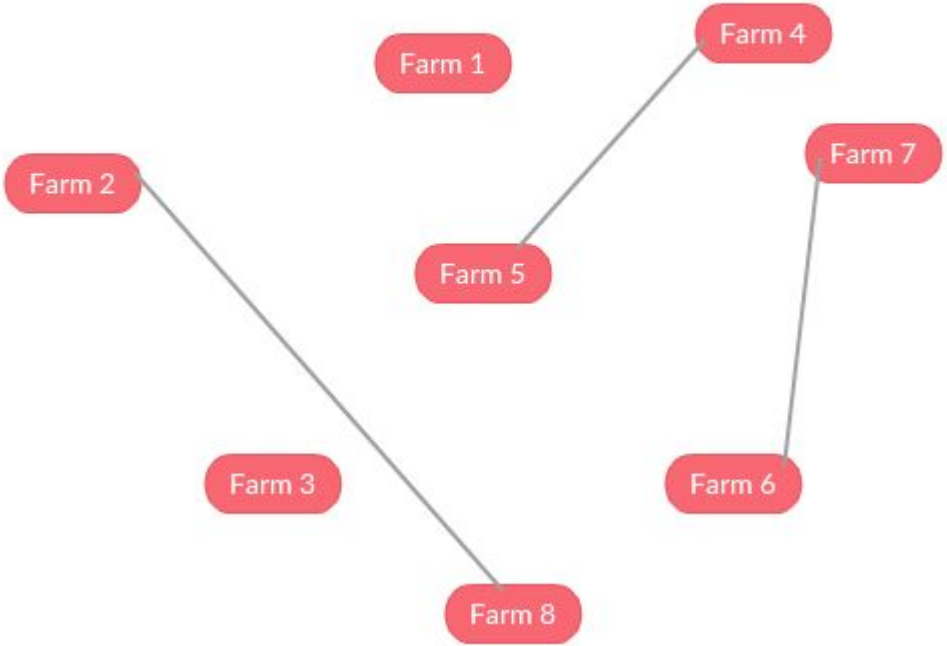
Graph based on temperature

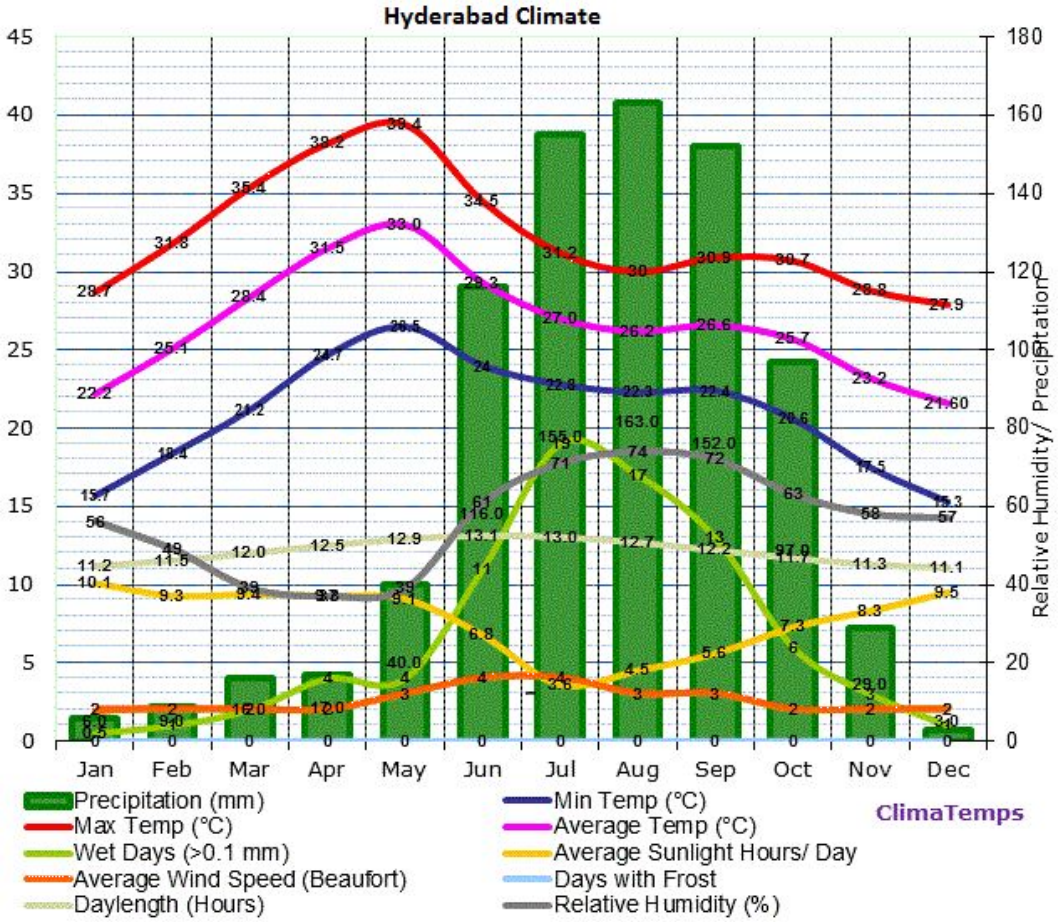


Graph based on rainfall



Graph based on crop species




Hyderabad Climate
Relative Humidity/ Precipitation
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Precipitation (mm)
Max Temp (°C)
Wet Days (>0.1 mm)
Average Wind Speed (Beaufort)
Daylength (Hours)
Min Temp (°C)
Average Temp (°C)
Average Sunlight Hours/ Day
Days with Frost
Relative Humidity (%)
ClimaTemps

Models

Graphical method

- Taking into account all the dependencies in the graphs, we model the system.

Spectral analysis

- Considering timely behaviour of the respective feature.

Topic modeling

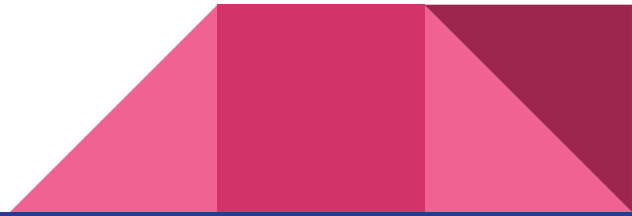
- Discovering the abstract symptoms using keywords.

Predicting the solution using Multi View Learning techniques.

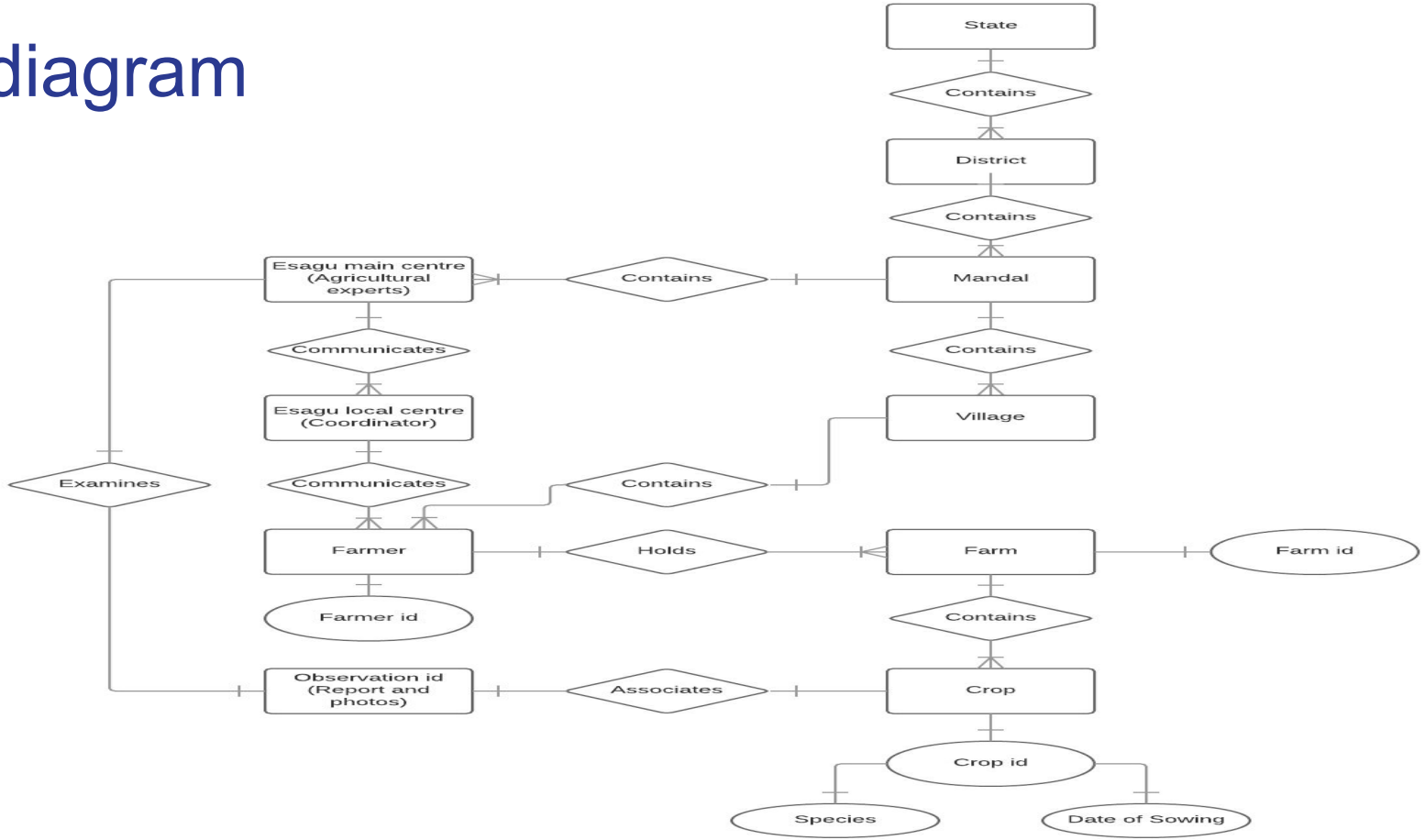
- Specific MVL solution will depend on the data.
- If the dimension is large, then we have to use Canonical Correlation Analysis.
- If the entire data is labeled, we can use multiview supervised learning.
- If we have both labeled and unlabeled data, then we have to use multiview semi supervised learning.
- Several such procedures like active learning, ensemble learning etc.
- Finally, a feedback to improve the accuracy of the prediction algorithm.

Progress

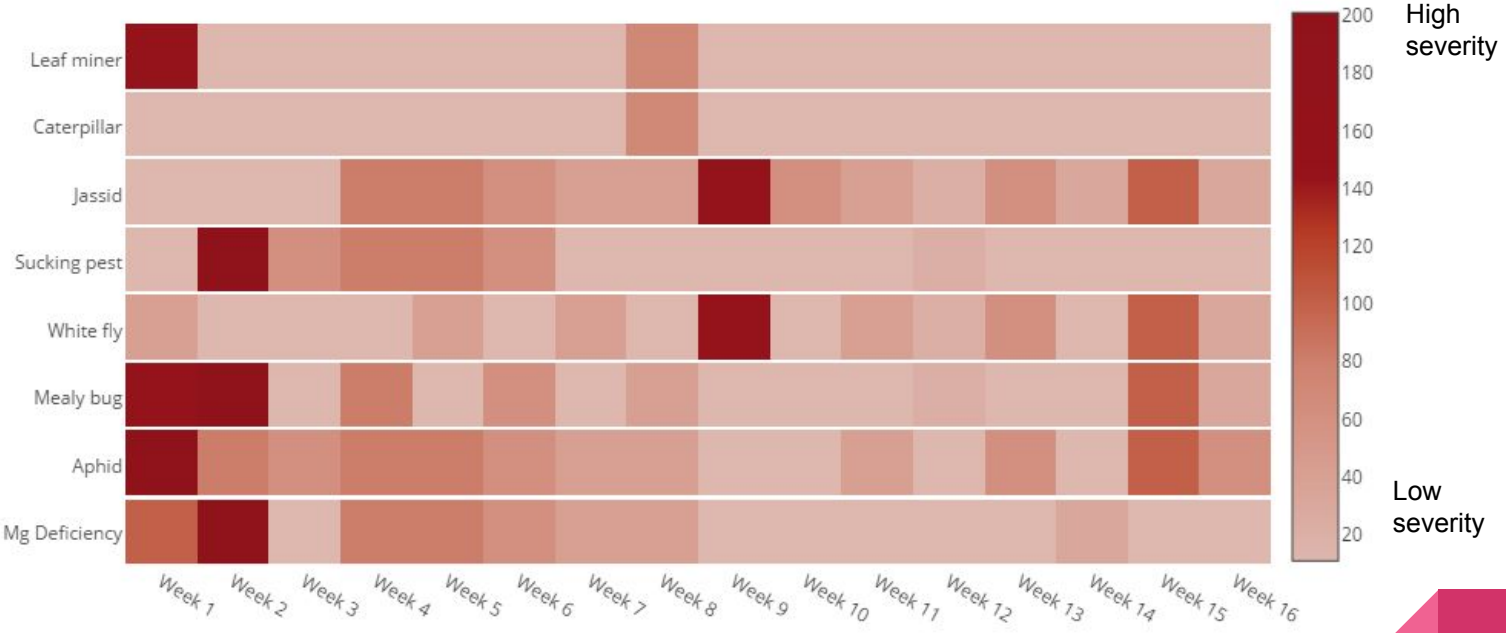
- First, we were given the data regarding a single cotton crop for one season.
- Heatmap visualization was performed on that data.
- Later, text analysis was performed on 5 year data regarding several crops across several villages.
- Visualization was done using several plots like histograms and pie charts.



ER diagram


State
Contains
District
Contains
Mandal
Contains
Village
Contains
Esagu main centre (Agricultural experts)
Communicates
Esagu local centre (Coordinator)
Communicates
Examines
Farmer
Farmer id
Holds
Farm
Farm id
Contains
Crop
Crop id
Species
Date of Sowing
Observation id (Report and photos)
Associates

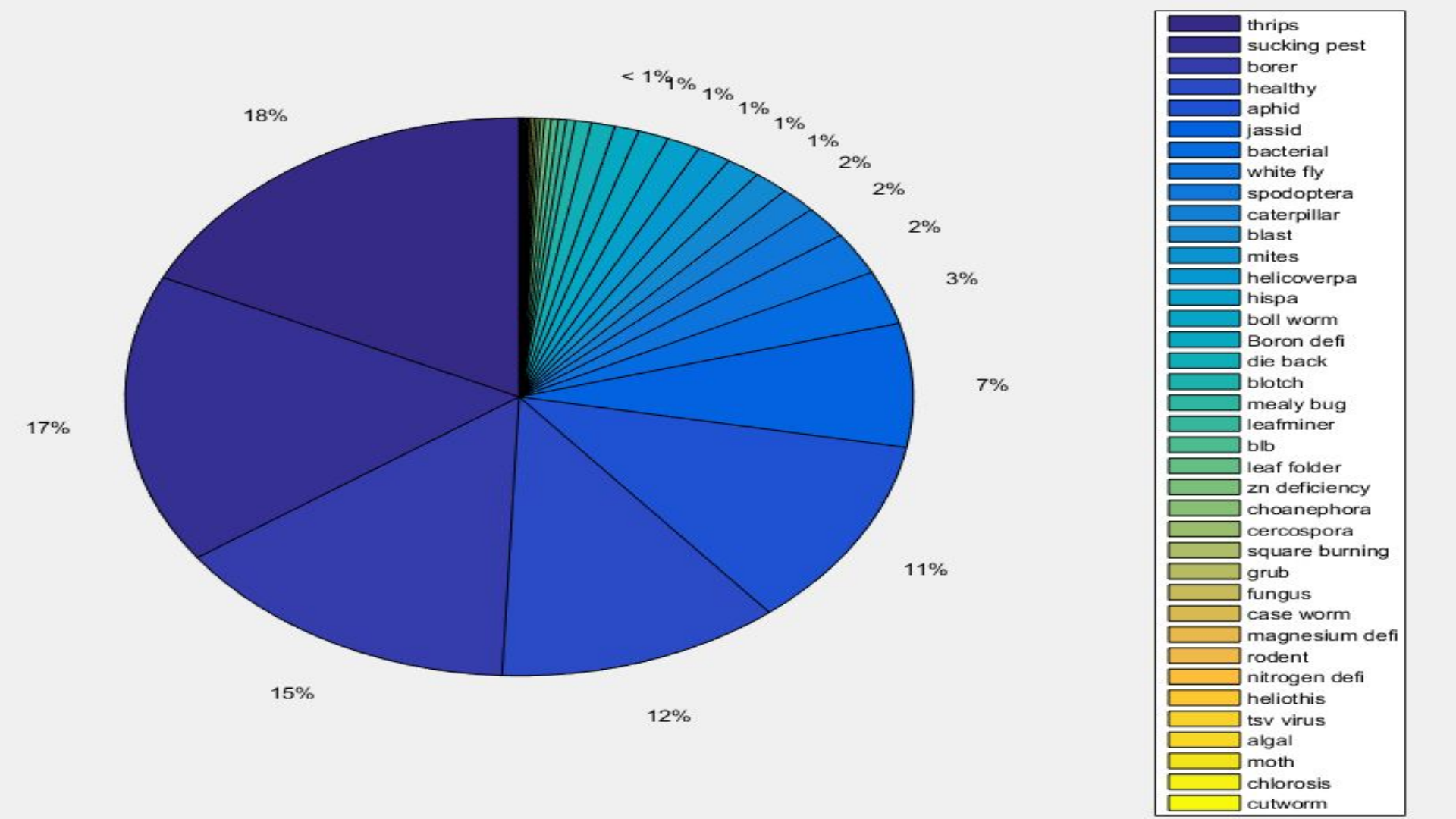
Dynamics of a single cotton crop (25-7-16 to 7-11-16)

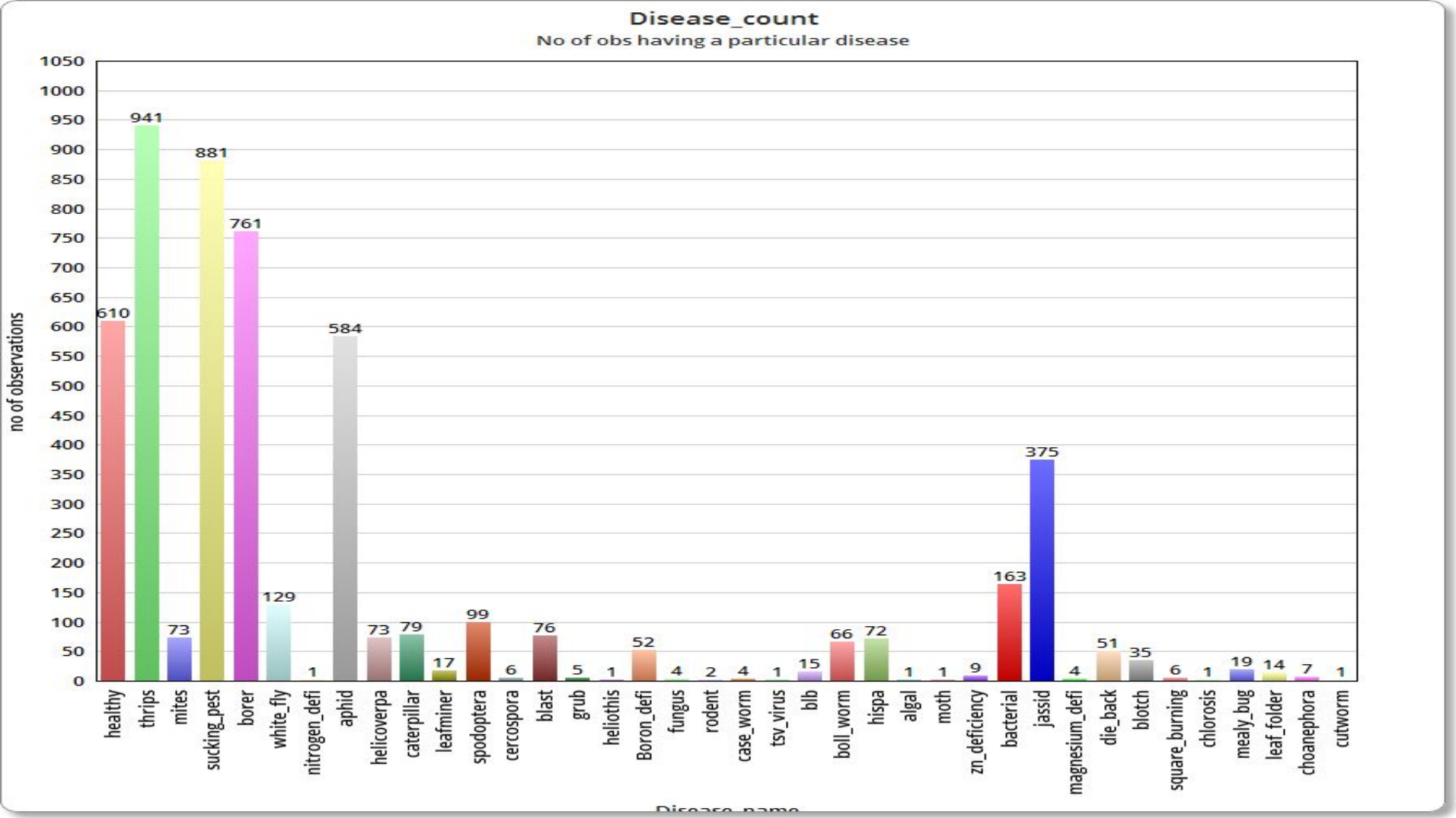


Visualization of provided data

Text analysis

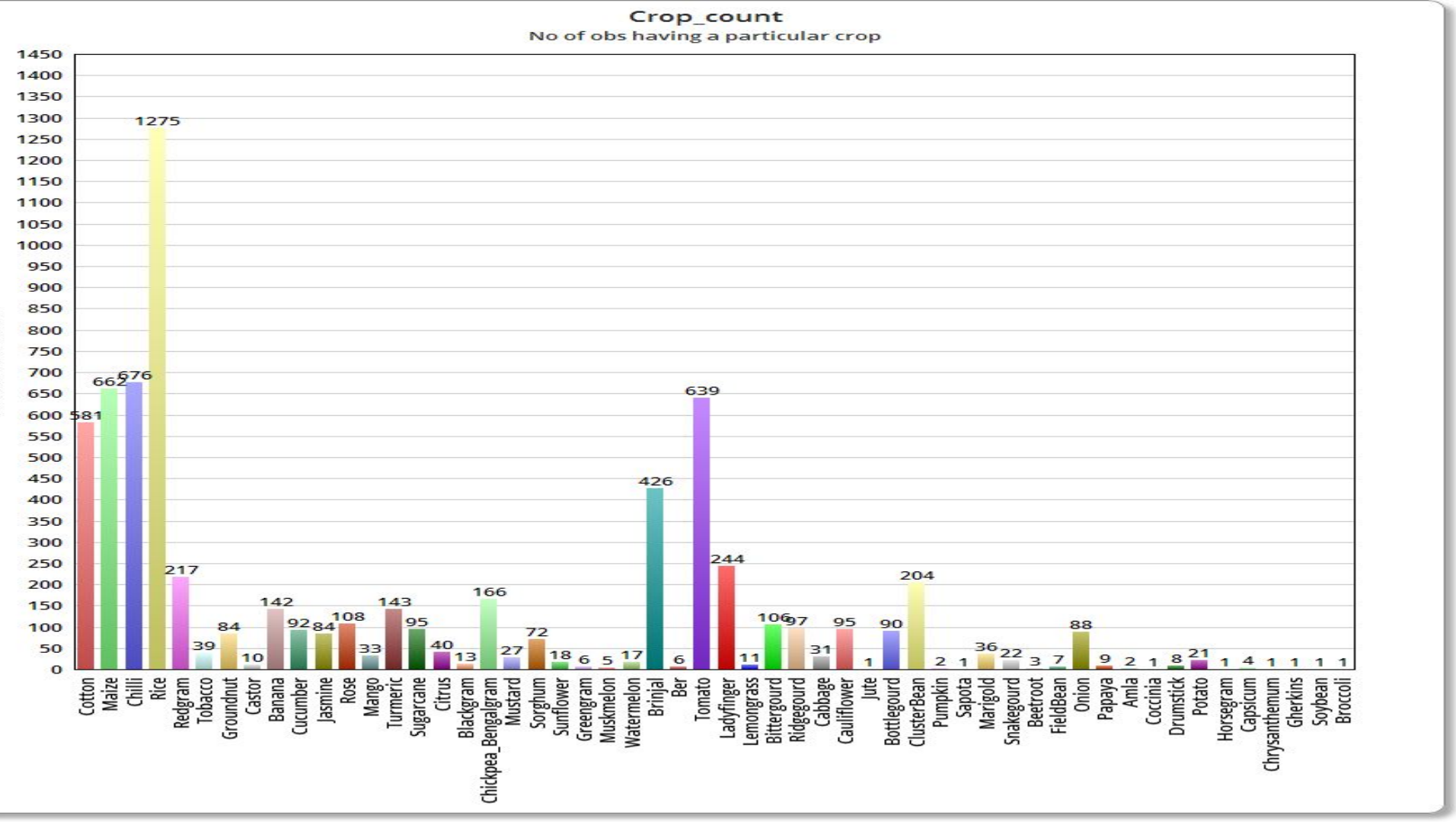
- Performed on data corresponding to several farms for 5 years (2012 to 2016).
- There were around 10,000 advices.
- The occurrences of different problems were counted by identifying corresponding keywords.
- Advices containing more than one problem would increment the count of respective problems

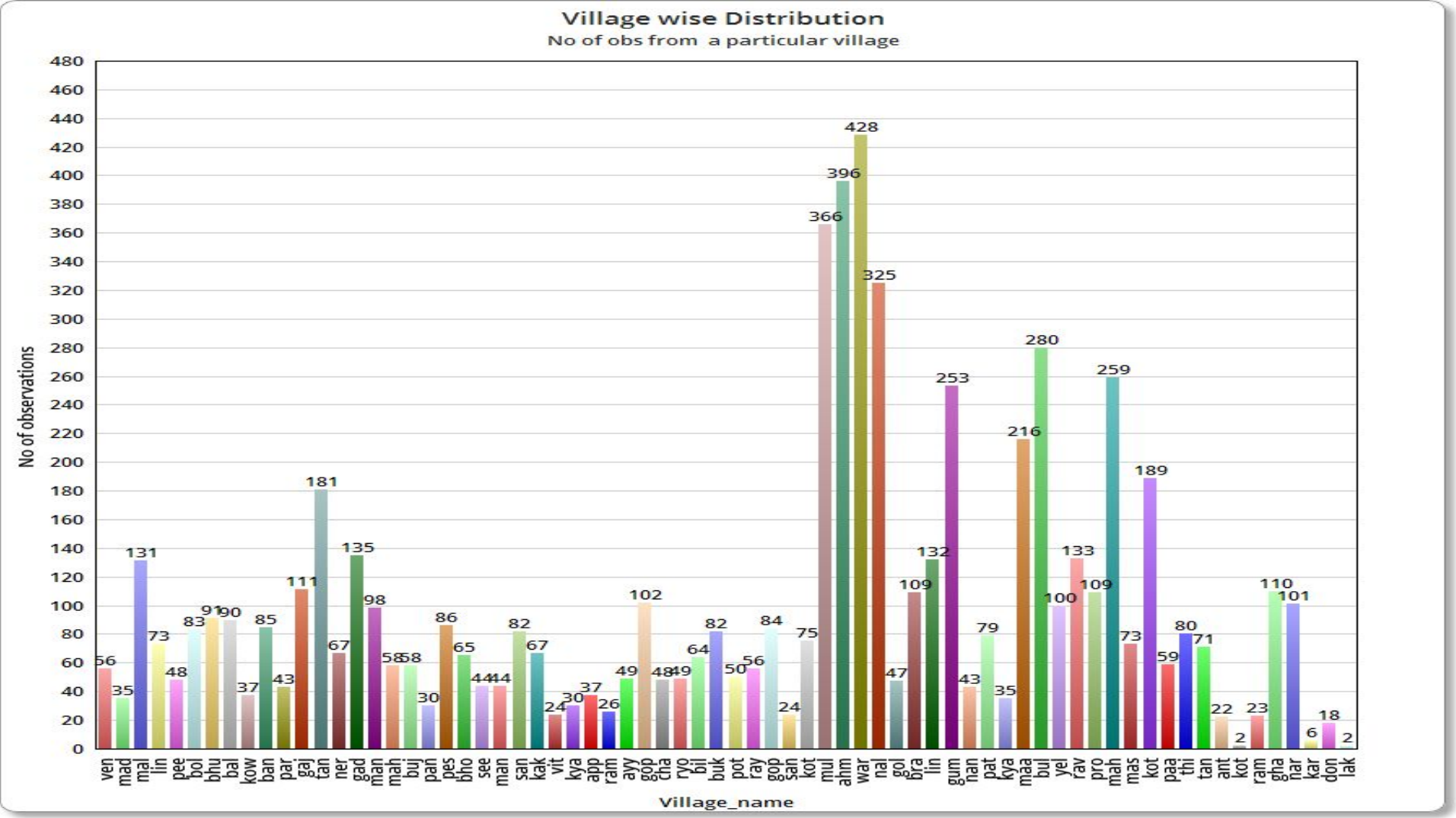

< 1%
1%
1%
1%
1%
1%
2%
2%
2%
3%
7%
11%
12%
15%
17%
18%
thrips
sucking pest
borer
healthy
aphid
jassid
bacterial
white fly
spodoptera
caterpillar
blast
mites
helicoverpa
hispa
boll worm
Boron defi
die back
blotch
mealy bug
leafminer
blb
leaf folder
zn deficiency
choanephora
cercospora
square burning
grub
fungus
case worm
magnesium defi
rodent
nitrogen defi
heliothis
tsv virus
algal
moth
chlorosis
cutworm


Disease_count
No of obs having a particular disease
no of observations
1050
1000
950
900
850
800
750
700
650
600
550
500
450
400
350
300
250
200
150
100
50
0
610
941
73
881
761
129
1
584
73
79
17
99
6
76
5
1
52
4
2
4
1
15
66
72
1
1
9
163
375
4
51
35
6
1
19
14
7
1
healthy
thrips
mites
sucking_pest
borer
white_fly
nitrogen_defi
aphid
helicoverpa
caterpillar
leafminer
spodoptera
cercospora
blast
grub
heliothis
Boron_defi
fungus
rodent
case_worm
tsv_virus
blb
boll_worm
hispa
algal
moth
zn_deficiency
bacterial
jassid
magnesium_defi
die_back
blotch
square_burning
chlorosis
mealy_bug
leaf_folder
choanephora
cutworm
Disease_name

But here, the entries do not sum up to 10,000. It is due to several reasons, some of which are :

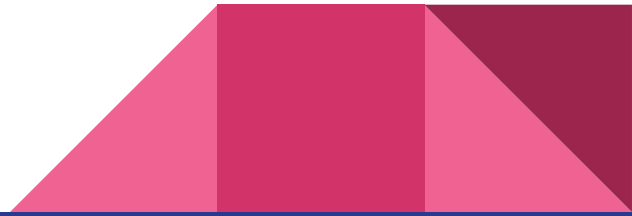
- Most of the advices are regarding fertilizers and do not contain the problem.
- Advices containing general preventive measures.


Crop_count
No of obs having a particular crop
1450
1400
1350
1300
1250
1200
1150
1100
1050
1000
950
900
850
800
750
700
650
600
550
500
450
400
350
300
250
200
150
100
50
0
581
662
676
1275
217
39
84
10
142
92
84
108
33
143
95
40
13
166
27
72
18
6
5
17
426
6
639
244
11
106
97
31
95
1
90
204
2
1
36
22
3
7
88
9
2
1
8
21
1
4
1
1
1
1
Cotton
Maize
Chilli
Rice
Redgram
Tobacco
Groundnut
Castor
Banana
Cucumber
Jasmine
Rose
Mango
Turmeric
Sugarcane
Citrus
Blackgram
Chickpea_Bengalgram
Mustard
Sorghum
Sunflower
Greengram
Muskmelon
Watermelon
Brinjal
Ber
Tomato
Ladyfinger
Lemongrass
Bittergourd
Ridgegourd
Cabbage
Cauliflower
Jute
Bottlegourd
ClusterBean
Pumpkin
Sapota
Marigold
Snakegourd
Beetroot
FieldBean
Onion
Papaya
Amla
Coccinia
Drumstick
Potato
Horsegram
Capsicum
Chrysanthemum
Gherkins
Soybean
Broccoli


Village wise Distribution
No of obs from a particular village
No of observations
Village_name
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
360
380
400
420
440
460
480
ven 56
mad 35
mal 131
iin 73
pee 48
bol 83
bhu 91
bal 90
kow 37
ban 85
par 43
gaj 111
tan 181
ner 67
gad 135
man 98
mah 58
buj 58
pan 30
pes 86
bho 65
see 44
man 44
san 82
kak 67
vit 24
kya 30
app 37
ram 26
ayy 49
gop 102
cha 48
ryo 49
bil 64
buk 82
pot 50
ray 56
gop 84
san 24
kot 75
mul 366
ahm 396
war 428
nal 325
gol 47
gra 109
iin 132
gum 253
nan 43
pat 79
kya 35
maa 216
bul 280
yel 100
rav 133
pro 109
mah 259
mas 73
kot 189
paa 59
thi 80
tan 71
ant 22
kot 2
ram 23
gha 110
nar 101
kar 6
don 18
lak 2

Timeline

- One week from now : Developing dashboard for interactive visualization of the data and prescriptive analytics service.
- One month from now : Model implementation.
- Next semester : recommender system.



References

- http://insait.in/AIPA2012/articles/009.pdf
- http://www.saravananraj.net/wp-content/uploads/2014/12/45_AFITA_ICT-for-Agricultural-extension_India.pdf
- http://insait.in/AIPA2012/articles/005.pdf
- https://www.researchgate.net/publication/233572675_Analysing_dynamics_of_crop_problems_by_applying_text_analysis_methods_on_farm_advisory_data_of_eSaguTM.