

Agriculture data analytics

—

- There are 54 different species of crops in the data.
- These crops are grown across 3 seasons :
 - Kharif (July-October)
 - Rabi (October-March)
 - Summer (March-June)
- For each of these 3 seasons, we find the set of probabilities for the occurrence of different diseases under various weather conditions separately.

For each season, we first fix the crop species. Then for each species of crop, we find the following probabilities:

- Probabilities of occurrence for all the 38 problems i.e, $p(\text{problem})$ for given species of crop.
- Probability of weather parameter being at certain value, given a particular problem i.e, $p(\text{weather parameter}/\text{problem})$. We estimate this by finding the values of all the weather parameters corresponding to the dates of problem occurrence.
- Probability of the crop being at certain age, given a particular problem i.e, $p(\text{age of crop}/\text{problem})$.

Once we find all these probabilities, we can estimate the probabilities of occurrence of each problem under given weather conditions and age of the crop, for given crop. For example, let us say the report we received is of the crop Maize.

- First, we look at the season in which it is being grown. Based on that, we select the corresponding set of probabilities.
- Then, we look at the age of the crop and all the weather parameters including temperature, humidity etc. on that day of observation.

Then, for each problem, we find this probability :

$$p(\text{problem}/[\text{temperature, humidity, age, other weather parameters}]) =$$

$$\frac{p([\text{temp, humidity, age, other parameters}]/\text{problem}) * p(\text{problem})}{p([\text{temp, humidity, age, other parameters}])}$$

$$= \frac{p(\text{temp}/\text{pb}) * p(\text{humidity}/\text{pb}) * p(\text{age}/\text{pb}) * p(\text{other}/\text{pb}) * p(\text{problem})}{p(\text{temp}) * p(\text{humidity}) * p(\text{age}) * p(\text{other parameters})}$$

$$= \frac{p(\text{temp}/\text{pb}) * p(\text{humidity}/\text{pb}) * p(\text{age}/\text{pb}) * p(\text{other}/\text{pb}) * p(\text{problem})}{p(\text{temp}) * p(\text{humidity}) * p(\text{age}) * p(\text{other parameters})}$$

$$= \frac{p(\text{temp}/\text{pb}) * p(\text{humidity}/\text{pb}) * p(\text{age}/\text{pb}) * p(\text{other}/\text{pb}) * p(\text{problem})}{p(\text{temp}) * p(\text{humidity}) * p(\text{age}) * p(\text{other parameters})}$$

(Assuming independence between different parameters)

We find the top 3 problems with highest probabilities and suggest them as the possible problems.

We already have the probabilities of occurrence of each problem and also the values of all the weather parameters on the dates of occurrence of all the problems. Using these values, we have to first estimate the probabilities p (weather parameter/problem) , taking into account the lack of observations for certain values. We have to fill up the probabilities for missing values by using interpolation or some other method.

We will try to come up with a method to take care of missing values by this weekend.

Thank you