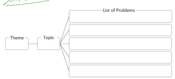
Problem Scoping

(i) It is a fact that we are surrounded by problems. (ii) They could be small or big, sometimes ignored or sometimes even critical. (iii) Many times we become so used to a problem that it becomes a part of our life. (iv) Identifying such a problem and having a vision to solve it, is what Problem Scoping is about.

Schema → To identify problem we'll follow schema.



Theme: Digital Literacy **Topics:** Online learning platforms, digital awareness, e-books, etc.
Theme: Health **Topics:** Medicinal Aid, Mobile Medications, Spreading of diseases, etc.
Theme: Entertainment **Topics:** Media, Virtual Gaming, Interactive AVs, Promotions etc.

Problem statement template

Our	[stakeholder(s)] Farmers, Fertilizer producers, Labours.	Who
has /have a problem that	[issue, problem, need] Determine what will be the best time for seeding or crop harvesting.	What
when / while	[context, situation] Decide the mature age for the crop & determine its time.	Where
An ideal solution would	[benefit of solution for them] Take the crop on time & supply against market demand on time	Why

Agriculture → Yield rates. Sowing & harvesting Patterns. → How might we help farmers determine the best times for seeding & sowing their crops?

4Ws Problem Canvas

Who?

✓ The "Who" block helps you in analysing the people getting affected directly or indirectly due to it. Under this, you find out who the 'Stakeholders' to this problem are and what you know about them. Stakeholders are the people who face this problem and would be benefitted with the solution.

Q. Who are Stakeholders?
Ans. Farmers, Fertilizer producers, labours & Tractor companies.

Q. What do we know about them?
Ans: These are some of the people worst affected by the problem & loses their money & time.

What?

Under the "What" block, you need to look into what you have on hand. At this stage, you need to determine the nature of the problem. What is the problem and how do you know that it is a problem? Under this block, you also gather evidence to prove that the problem you have selected actually exists. Newspaper articles, Media, announcements, etc are some examples.

What?

- What is the problem?
  Determine what will be the best time for seeding or crop harvesting.
- How do you know that it is a problem? (Is there any evidence?)
  Seeding at improper position & imperfect time leads to the wastage of money & time.

Where?

Now that you know who is associated with the problem and what the problem actually is; you need to focus on the context/situation/location of the problem. This block will help you look into the situation in which the problem arises, the context of it, and the locations where it is prominent.

Where?

- What is the context/ situation the stakeholders experience the problem?
  Decide the mature age for the crop & determine its time.
- Where is the problem located?
  In the field, certain parts of country.

Why?

Why?

You have finally listed down all the major elements that affect the problem directly. Now it is convenient to understand who the people that would be benefitted by the solution are; what is to be solved; and where will the solution be deployed. These three canvases now become the base of why you want to solve this problem. Thus in the "Why" canvas, think about the benefits which the stakeholders would get from the solution and how would it benefit them as well as the society.

Why?

- Why will this solution be of value to the stakeholders?
  This solution will help us to determine the exact time for sowing & harvesting of crops, thus leading to max yield.
- How will the solution improve their situation?
  They could maximize their profit.

Data Acquisition

✓ It is the stage in which we acquire data for the project.

What is Data?

- Data can be defined as a representation of facts or instructions about some entity (school, sports, business, animals, students etc) that can be processed or communication by human or machine.
- Data is a collection of facts, such as numbers, words, pictures, audio clips, videos, map, measurements, observations or even just description of things.
- Data may be represented with the help of characters such as alphabets (A-Z, a-z), digits (0-9) or special characters (+, -, /, *, <, >, = etc)

Data → Structured Data | Unstructured Data

Structured Data: Highly organised & Easily analysed

Name	Age	Class

Structured Data:

→ Structured data is categorized as quantitative data.
→ It's the type of data most of us work with every day.
→ Structured data has predefined data types & format so that it fits well in the column/fields of database or spreadsheet.
→ Highly organised & easily analysed.
→ Data is stored in relational databases (storing in form of tables) or spreadsheets (like Excel sheets)
→ Examples of structured data are Name, age, address, etc.

Unstructured Data:

→ qualitative data
→ Can't be processed & analysed using conventional RDBMS methods
→ Unstructured data is difficult to deconstruct because it has no predefined model, meaning it can't be organized in relational databases.
→ NOSQL databases for managing unstructured data.
→ Example of unstructure data :- video, audio, satellite imagery, social media activity.

Datasets

→ A Data set is just a set or Collection of data.

→ This set is normally represented in tabulous pattern.

→ Every column describes a particular variable.

→ Each row corresponds to a given member of the data set.

Features

Datasets

Name of Student	Attendance in %	Gender	Participation in Sports	Total marks obtained (%)
A	76	Male	N	92
B	62	Female	Y	88
C	63	Male	Y	93
⋮	⋮	⋮	⋮	⋮
F	93	Male	Y	76

80% Training Dataset

Test Dataset

Feeding data to machine for training

✓ Training Dataset

✓ Test Dataset.

✓ Features. ← Col's selected for prediction purpose are known as features

80%
20%

Data features

→ A measurable piece of data that can be used for analysis.

→ In CSV & Excel files they could be seen as columns.
↑
Comma separated value

→ Features are also sometimes referred to as "Variables" or "attributes".

→ Depending on what we're trying to analyze, the features we include in our dataset can vary widely.

Acquiring Data (Sources)

Surveys	Web Scraping	Sensors
Cameras	Observations	API(Application Program Interface)

Surveys

✓ A research method used for collecting data from a predefined group of respondents to gain information & insights into various topics of interest.

Cameras

→ A camera captures a visual image.
→ Could be used to collect data for CV projects
→ A device for recording visual images in the form of photo, film or video.

Web Scrapping

✓ Web scrapping is the process of collecting structured web data in an automated fashion, It's also called web extraction.

✓ Use cases of web scrapping include price monitoring, price intelligence, news monitoring, lead generation, & market research among many others.

Observations

✓ Some data we can acquire through monitoring & close inspection.

Sensors

✓ A device which detects or measures a indicated property & records, indicates, or otherwise responds to it.

✓ Eg - Temperature Sensors, Humidity Sensors, Pressure Sensors, Proximity Sensors, level Sensors, Accelerometers.
↑ linear change

Gyroscope
↑ angular ↑ Measures Rate change

Infrared sensors — (i) Ex - Cash counting m/c.
(ii) In Mall Infrared sensors:
Count no of people in
&
count no of people out.

API (Application Program Interface).

→ It is a s/w intermediatary that allows two applications to talk to each other.
backend + frontend.

Acquiring Data from Reliable Sources.

✓ Internet, Random websites.
↑
Data which we collect is open-sourced & not someone's Property.

✓ open source websites hosted by the Govt.

✓ Some other sources: Kaggle Dataset, Amazon Dataset, Yahoo webscope.

System Maps :-

→ helps us to find relationship b/w different elements of the problem which we have scoped.

→ To understand complex issues with multiple factors that affect each other.

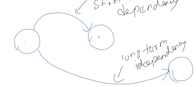
→ Every element is interconnected.

→ One can easily define a relationship among different elements which come under a system

In a system map

Circles
For representing data features →

Arrows
↑
For Interconnecting features



loops

Forming a continuous circuit

+ and - sign

↑ marks ∝ hard work you
↑ marks ∝ 1 / Amount of time you was

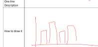
+ - direct proportional.
- - Inverse proportion

Data exploration

✓ Goal is to learn about characteristics & potential problems of a data set without the need to formulate assumptions about the data beforehand.

✓ In statistics, data exploration is often referred as "Exploratory data analytics"

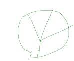
To do two things:
i) Cleaning the data.
ii) We'll understand Graphs.

Data Visualisation Technique 1	
Name of the Representation	Bar graph or column graph
One-line Description	
How to draw it	
Suitable for which data type?	Comparisons & functions.

The classic Bar chart uses either horizontal → or Vertical bars to show discrete or numerical comparisons across categories.

Pie chart:

Pie chart help show proportions & percentages between categories, by dividing a circle into proportional segment. Each arch length represents a proportion of each category, while the full circle represents the total sum of all data, equal to 100%



Comparisons, Part-to-a whole & proportions.

Histogram

– A Histogram visualizes the distribution of data over a continuous interval or certain time period. Each bar in a histogram represents the tabulated frequency at each interval/bin.

Histograms are great when we would like to show the distribution of the data we are working with.

Func) Comparision, Data over Time, Distributions Patterns, Range

Scatterplot